

Evaluating Social Impact Bonds

Developing a generic evaluation tool for Social Impact Bonds



Evaluating Social Impact Bonds

Developing a generic evaluation tool for Social Impact Bonds

by

Björn Vennema

to obtain the degree of Master of Science
at the Delft University of Technology,

Student number: 1510479
Project duration: Februari 1, 2015 – February 18, 2016
Thesis committee: Prof. dr. M. S. van Geenhuizen, TU Delft
Dr. Z. Roosenboom-Kwee, TU Delft
Dr. H. G. van der Voort, TU Delft

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

*My dream is to find individuals
who take financial resources and
convert them into changing the
world in the most positive ways.*

JACQUELINE NOVOGRATZ

Preface

When I first started working on Social Impact Bonds, late 2012, I had no idea what they are and why they were used. I instantly thought that it was something interesting for my employer at the time, ABN AMRO, albeit for completely the wrong reasons. A year's work later and I was part of a very small team that launched the first SIB in Continental Europe. An exciting but small step towards much more to come. I have continued to work on SIBs and was granted the opportunity to work as an intern at the birthplace of SIBs and a worldwide front runner of social entrepreneurship, Social Finance UK. This experience has formed me in ways I couldn't believe, but has also inspired this thesis.

As I've learned, SIBs are started from a practical ideology and brings people from all sorts together who all want to make the world a slightly better place for everyone. SIBs were, and are, designed from the idea of solving a problem and has delivered great results. In a fast moving world it is sometimes important though, to stand still and overthink whether you are still doing the right thing. That is precisely what I've tried to accomplish with this thesis. Not only standing myself, but also by designing an evaluation tool that allows all stakeholders in SIBs to do the same on a much more frequent base. The world is evolving continuously and while it is a great step to bring different stakeholders together in order to help those who need it, it is even better if we can ensure that at all times, help is offered in the best way we can think of.

Everyone who knows me, also knows that writing this thesis hasn't been a smooth ride. Not just the practicalities of it, but just getting through it has proven a challenge for me personally. I feel proud of the result however and hope that I have truly attributed something to the development of SIBs. While for some, evaluation might not sound like the most fun thing to do, I think it is probably the most important aspect of everything that is in development. From someone as a person, to SIBs.

There are too many to thank for their help in completing this thesis, so please accept my apologies if you are not on here, I will make it up to you. Obviously I would like to thank my parents, Fulco en Wilma, for raising me and always giving me the opportunities to explore my abilities, encouraging me to be the best I can and not getting frustrated for the extra years that I had to spend at university. My girlfriend Evelien for being a great support and putting up with me being in a different country 9 months, frequently staying too late in the office to even call or Skype. Everybody at Social Finance UK for giving me one of the best experiences of my life, which I will never forget, but especially Fiona Miller Smith and Jane Newman for allowing me to join the team, I hope to be back at some point. My thesis committee Prof. Van Geenhuizen, Dr. Roosenboom-Kwee and Dr. Van der Voort, all have stepped in on a different point in time, but have been incredibly supportive and thanks to your stepping in, I have actually been able to finish this thing. Finally a special thanks to Ruben Koekoek, for introducing me to SIBs, teaching me loads and obviously participating in this thesis as an interviewee. I am proud to have you as a friend now.

An additional thanks to all who have participated in an interview with me throughout the past months, providing input and insights that have make me think further about SIBs and evaluations: Paul Hopkins, Antony Ross, Janette Powell, Bart van Hulst, Udo Pesch, Bert van Enserink, Leo van Loon, Tjalling de Vries and Christian van Esch.

On to the next adventure, launching more and better Social Impact Bonds and learning as much as possible through evaluation.

*Björn Vennema
Amsterdam, February 2016*

Summary

This thesis looks into the development of a generic evaluation tool for Social Impact Bonds (SIBs). As a relatively new concept, SIBs are bringing change on important aspects of several sectors. The global financial crisis has not only exposed weaknesses in the financial sector, but also once again put the spotlight on the inequality that is existent in many parts of the world. Since, there has been growing interest in sustainability and realising social impact through many ways. This also brings up additional complexities though. Questions as, who is paying the bill and how do we measure social impact that has been realised become more important than ever, but are also much harder to answer than the comparison of the profits of traditional companies. SIBs provide a solution of bringing stakeholders from both the public and private sector to not only achieve, but also scale the social impact they deliver. So what are they? This thesis takes the definition that has been stated by Gustafsson et al: *a Social Impact Bond is a model in which private investors put up capital to fund a social intervention and governments repay the investors only if an agreed-upon outcome is achieved* [33][p.4].

The launch of the first SIB took place in 2009, when the Peterborough SIB was initiated. Now in 2016, the total number of SIBs is closing in on 50 and more SIBs are being launched each year. With inequality and many social issues remaining significant problems of society, new ways are needed to improve the results of charities and other social interventions. At the same time, there are many ways in which money could be used better in order to provide value for society. SIBs are growing in a time where more and more organisations are striving to look for more responsible options, for example the European Union has launched the initiative of Responsible Research & Innovation. Aiming to enable more responsible innovation, SIBs are exactly such, an innovative partnership to create more social impact.

SIBs have arisen as a practical solution though and is being launched to solve many different social issues. Sometimes it is good to stop and think about what you have been doing and whether that can be improved. The SIBs clearly need that after those first 6 years of development. But not only to think about the concept of SIBs, but also within individual SIBs, evaluation should be an important aspect to learn from the current SIBs and move forward in a more effective way. This thesis sets out to design a generic evaluation tool that is applicable to all SIBs. This tool should create evaluation that is both valid and fair, provides learnings for SIBs and provides results that can be compared amongst SIBs, even when they are working with different topics. The main research question based on this objective has been *What are the design characteristics of a tested evaluation tool for Social Impact Bonds*.

Initially, the main criteria of an evaluation tool are developed. These criteria are used to provide input for the evaluation tool and are developed based on theoretical research and interviews with researchers from various fields of study. The criteria should take into account the complexity of the SIB and the different incentives for all stakeholders as identified in chapter 2, but also evaluate the whole system of SIBs. Therefore a total of 8 criteria have been developed.

- Both quantitative and qualitative evaluation
- Separate evaluations that have service users and service providers as units of analysis
- Outcome indicators should focus on objective and agreed upon criteria
- Monetisation should not be part of the evaluation
- Effects on society and on people's well-being should be evaluated
- Evaluation uses baseline data and ex-ante, ex-post evaluation
- Monitoring and subsequent discussion of the monitoring results should take place at least every three months and in between each cohort of a project (when applicable)
- Evaluation should be performed by a third party

The costs of a SIB can clearly become an issue for the evaluation, as high costs might make the SIB too expensive. Therefore, these criteria have focused on enabling a tool that has high validity as well as practical achievability.

To enable a sound evaluation of the impact of SIBs on the well-being of the participants of the services, a deeper investigation has been performed on the Capabilities Approach. This approach has been used in several instances to evaluate the well-being of individuals and is most prominently used as the basic for the Human Development Index as used by the United Nations [87]. It clearly has weaknesses on its operationalisation, as this remains a sensitive project for something as delicate as the well-being of people. Using other studies, a questionnaire has been developed that would evaluate the well-being of participants based on all ten central capabilities as defined for the Capabilities Approach by Martha Nussbaum [47].

The questionnaire based on the Capabilities Approach is part of a broader initial design of the evaluation tool. With two main elements, evaluation and monitoring, several aspects have been developed for an initial design that would be later tested.

Firstly, the evaluation element has been split out in the evaluation of the outcomes of a SIB, evaluation the impact on participants, evaluation the service provider's processes and finally the SIB's processes. The only aspect that can't be a fully generic aspect is that of evaluating the outcomes, as the outcomes and its indicators will be different for each SIB. Nonetheless, the steps towards correctly evaluating those outcomes have been set out. The first step is to determine the criteria for baseline data. While, from a validity perspective, setting up a randomised control trial would deliver more valid evaluations, this proves to be both ethically and practically troublesome. The second step is to collect data at the start of the intervention for baseline and participating service users. While the latter is essential, the ability to also collect baseline data both before and after the intervention highly improves the possibility to exclude more of the exogenous effects on a SIB. The third step is a repetition of the second step, but then after the intervention has finished for the participants. Finally the collected data would be analysed and corrected for any exogenous possible to be distinguished from the data, to provide the actual effect the SIB has had on the outcome indicators.

The second aspect, assessing the impact on participants is the questionnaire based on the Capabilities Approach as discussed earlier. Finally, it is established that it is important to evaluate the processes of a SIB. This will not only evaluate the outcomes, but another important part of evaluation is to establish learnings that can be implemented for future improvements. Therefore an additional questionnaire was designed for the participants on their experiences with the processes of the service provider. Finally a questionnaire was also prepared for the primary stakeholders of the SIB, to evaluate the system of the SIB itself.

Another important element of the evaluation tool is that of monitoring the SIB during its activities. It is proposed that a SIB is monitored at least every three months on two main aspects. First, the achieved outcomes and the ongoing inputs and outputs, second on the service provider's processes. The outcomes of the monitoring should be discussed by all stakeholders and lead to negotiations about the next steps. The combination of the monitoring results and the discussions will provide insights for the SIB that can lead to adaptive decision making to continuously try to improve the social impact of a SIB.

Finally, the governance of the tool is very important. The most important role is obviously played by the evaluator. The evaluator should as independent as an organisation can be in a marketplace. While dependencies cannot be ruled out completely, the evaluator should stand to gain no benefits from all outcomes that the evaluation may provide. It also should make the evaluation transparent and publicly available in order to reduce the risk of bias in the evaluation. While all stakeholders play a role in the evaluation of the SIB's processes, they should give the evaluator every opportunity to access the needed information without any barriers. This will mean that any legal and privacy issues on a third party accessing personal or governmental data should be sorted out before the start of the SIB.

This initial design was tested by interviewing several active stakeholders of SIBs as well as looking at a case study of all SIBs that are currently live.

The interviews provided interesting insights into the perspectives of the stakeholders of SIBs to evaluation. It became clear that all stakeholders highly valued valid evaluation of SIBs and saw it as a crucial aspect to a successful SIB. There also was a focus on the outcome indicators and the evaluation of the financial returns that are related to the SIB. This is seen as the primary focus of the evaluation. Especially the aspect of evaluating the impact on the participants was seen as an important addition

to the evaluation of SIBs, although for varied reasons. While the government saw it as something to use to evaluate the unintended outcomes, the service provider was more concerned with the fact that this might make the actual impact of a service more clear. The evaluation of the processes of the service provider was mostly criticised though, as it was seen as not adding value to the SIB or the evaluation of the SIB. Even the service provider acknowledged that this was something that was probably the responsibility of the service provider anyway. Finally, all parties also appeared to agree that independent and objective evaluation is necessary, not only for the validity of the evaluation, but also to create fairness among all stakeholders.

47 SIBs have been observed in the case study, looking predominantly at their set-up, the capabilities they target and the evaluation methodologies that are being used. As only few SIBs had finished at the time of writing this thesis and no clear outcomes have been published, it was not possible to relate those to the results of a SIB. Based on all analysis, it has become clear that the SIBs appear to target a limited set of capabilities with their service. The most targeted capabilities are Affiliation, through improved employment, and Senses, Imagination and Thought, through improvements in educational outcomes.

Mainly through the three large funds for SIBs in the UK, the most used evaluation methodology has been that of the tariff model, where outcome payments are based on a set of predetermined valuations of outcomes. Even though that methodology probably scores highest on practical achievability and is therefore useful to realise scale, this thesis prefers the second most used methodology of matched comparison evaluation. This uses baseline data and combines relatively high practical achievability with the required validity.

When all SIBs are observed and compared with the basic SIB system, three main typologies can be identified for the SIBs. SIBs where a participant consists of multiple people, i.e. family welfare, SIBs that target social issues that are primarily individual, such as improving the employment of individuals. The third typology is that of SIBs that target social issues that are primarily broader societal issues, such the reduction of recidivism. Even though these typologies provide useful insights, the SIB systems do not provide valuable additional insights for the evaluation tool designed.

The types of analysis also partly confirm the views of the interviewees. Most SIBs mainly focus on the evaluation on the outcome indicators and the financial components related to those. Monitoring is also an aspect that clearly comes across in several SIBs as well as the interviewees. Monitoring adds valuable insights to the stakeholders of SIBs and allows for adaptive decision making to make changes to the programme where possible. Although interviewees described making significant alterations as troublesome because of standing agreements, incremental improvements to the delivery of the SIB are nevertheless a key point of the monitoring tool. It can be seen in current SIBs that monitoring has been actively used to make such decisions, even when they have been as drastically as deciding to stop a SIB.

Based on the analysis of the interviews and the SIB case study, the evaluation tool was redesigned, leading to the final design as presented in Appendix I of this thesis. While some aspects have remained unchanged from the initial design, two aspects were changed significantly. Most rigorously the evaluation aspect of evaluating the service provider's processes was eliminated for the final design. Even though the processes of the service provider are seen as critical for delivering social impact, it is also something that can rightfully be argued to be the responsibility of a social entrepreneurs. Also, it would increase costs, while no stakeholder sees additional value to this aspect, making it unlikely to be implemented.

The other major change was in the questionnaire to evaluate the impact on participants. Based on the capabilities that were identified as mainly targeted by SIBs, a new questionnaire was designed that is both shorter and more targeted. Thus leading to a less expensive evaluation, but also increasing the response rates and quality by making the questionnaire better accessible for participants of the SIB. This questionnaire should still be given to each individual of the SIB to get an actual view of the number of people that are seen as improving certain capabilities.

No change were made to the governance of the SIB, where transparency and objectivity for the evaluator remain core aspects. While stakeholders themselves emphasised the need for trust in the competence of all stakeholders, the evaluator should clearly evaluate the roles, responsibilities and actions of all stakeholders. This would not only provide learnings for future SIBs, but also give insights in the biases and incentives that played a role during the SIB.

The final design presented by this thesis lays the groundwork for more coherent and valid evaluation for SIBs. When used correctly, this could provide more insights into the processes and outcomes of

SIBs and help innovate and scale the concept of SIBs further. Future research, but also implementation should focus on using this tool on new SIBs from the start and making those evaluations accessible. Only when used in practice is when this tool can provide additional impact to SIBs. Research should therefore focus on that implementation and the improvements it can find from it, both for SIBs as for this evaluation tool.

Contents

List of Figures	xi
List of Tables	xiii
1 Introduction	1
1.1 Introducing Social Impact Bonds	1
1.1.1 SIB System	3
1.1.2 Relevance	5
1.2 Research objective and questions	7
1.3 Structure of the thesis	8
2 Social Impact Bonds	11
2.1 Actor analysis	11
2.1.1 The commissioner	12
2.1.2 The service provider	13
2.1.3 The service user	14
2.1.4 The investor	14
2.1.5 The intermediary	15
2.1.6 The evaluator	15
2.1.7 Conclusions of the actor analysis	16
2.2 Trends that drive implementation of SIBs	16
2.2.1 Impact Investing	16
2.2.2 Payment by Results	16
2.2.3 Future?	17
3 Design of an evaluation tool	19
3.1 What is evaluation and why is it needed?	19
3.2 Criteria for the design of a tool to evaluate Social Impact Bonds	20
3.2.1 When should the evaluation take place	21
3.2.2 The evaluator	21
3.2.3 Validity of the evaluation tool	22
3.2.4 Practical achievability of the evaluation tool	22
3.3 Design methodology	23
3.3.1 What to evaluate and which indicators to use	23
3.3.2 Data availability and requirements	24
3.3.3 Methodology	25
3.3.4 Data collection	25
3.3.5 Data analysis	26
3.4 Conclusions of the criteria for an evaluation tool	27
4 Capabilities	29
4.1 The Capabilities Approach	29
4.1.1 Central capabilities	30
4.2 Critique on the Capabilities Approach	33
4.3 Operationalising individual capabilities	34
4.4 Measuring well-being	37
4.5 Conclusions of the Capabilities Approach	37

5	Characteristics of a generic evaluation tool for Social Impact Bonds	39
5.1	Evaluation.	39
5.1.1	Evaluating outcomes	39
5.1.2	Evaluating impact on participants	40
5.1.3	Evaluating the service provider's processes	40
5.1.4	Evaluating the SIB's processes.	41
5.2	Monitoring	42
5.2.1	Monitoring inputs and outputs	42
5.2.2	Monitoring service provider's processes	43
5.3	Governance of the tool	43
5.4	Conclusions of the initial design	44
6	SIB Analysis	47
6.1	Testing the evaluation tool on current stakeholders.	47
6.1.1	Overall interview responses	48
6.1.2	Criteria for the evaluation of Social Impact Bonds	48
6.1.3	Monitoring	50
6.1.4	Evaluation.	50
6.1.5	Conclusions from interviews	50
6.2	SIB case study	52
6.2.1	A high-level overview.	52
6.2.2	Workforce Development	53
6.2.3	Education and Early Years.	61
6.2.4	Housing/Homelessness	63
6.2.5	Child and Family Welfare	68
6.2.6	Criminal Justice	72
6.2.7	Health	74
6.2.8	Adults with Complex Needs	75
6.2.9	Conclusions on SIB case study.	75
6.3	Conclusions SIB Analysis	82
7	Redesign of the evaluation tool	85
7.1	Evaluation.	85
7.1.1	Evaluating outcomes	85
7.1.2	Evaluating impact on participants.	86
7.1.3	Qualitative evaluation	96
7.2	Monitoring	97
7.3	Governance	98
7.4	Conclusions of redesign.	98
8	Conclusions	99
8.1	Recommendations	102
A	Evaluation tool	105
B	Interview responses	113
	Bibliography	117

List of Figures

1	Cover photo by Dom Pates [52]	i
1.1	The general structure of a Social Impact Bond as proposed by McKinsey in 2012 [71, p.14]	2
1.2	Four criteria to make SIBs successful [33, p.5]	3
1.3	A simple example of the system of a SIB	4
1.4	A flowdiagram of this thesis	9
3.1	An example of the causal model between education and employment	23
6.1	An example system of the first typology	79
6.2	An example system of the second typology	80
6.3	An example system of the third typology	81

List of Tables

5.1	Overview of the evaluation tool of SIBs	45
6.1	Overview of the interviewees for this Thesis	47
6.2	Top ten responses	48
6.3	Rate card Innovation Fund	54
6.4	Rate card Youth Engagement Fund	59
6.5	Rate card of Fair Chance Fund	64
6.6	Capabilities that are aimed to be improved through existing SIBs	77
6.7	Evaluation methodologies on existing SIBs	78
B.1	All responses in interviews related to evaluation, codified and summed	114

Introduction

Traditional businesses has long focused primarily on one thing, financial returns. Even in 2016, the majority does. But there is also a growing group of entrepreneurs and investors that are becoming less focused on financial returns and are more interested in the social returns. They are interested in delivering social impact and doing good for society. This trend can be seen across industries and across the different types of actors. Large corporations have set up corporate social responsibility policies, which strive for improved sustainability. But growing numbers of start-ups can be seen as social enterprises, which are in their essence striving to achieve social impact. Finally, there is also a growth of the impact investment sector, where investments are done with the aim to realise both financial and social return. Since 2009, a concept is used that actually brings actors from totally different perspectives together to achieve that social impact, Social Impact Bonds. In this, government, investors and social entrepreneurs collaborate to realise impact for those who need it.

The increasing engagement with sustainability and social impact, but especially when such different stakeholders are involved in a single concept as with SIBs, also creates extra complexities. Social impact is something that is much harder to measure or compare than profit. While we can compare the profit of companies like Facebook and Procter & Gamble to see how they do financially, which largely has impact on how shareholder view the company. Comparing their impact on society is much harder to compare however, as they have such different types of impact. Measuring social impact therefore is important, but often cannot tell whether one project is better than another. With the growing involvement of different stakeholders, measuring social impact becomes increasingly important. Different stakeholders often means different incentives within a project which require well established tools to come to objective measurement and evaluation of companies or projects, but also require tools that can be comparable amongst completely different projects.

This thesis will look into the development of such an evaluation tool that is both generic as measures the essence of individual companies/projects. In the development of that tool, a relatively new product for involving different stakeholders in realising social impact will be used as a case study, Social Impact Bonds (SIBs). Before the case of SIBs is observed, the design process will be introduced and analysed. The design of an evaluation process will follow a number of standard steps to come to an initial evaluation tool. The first step is to decide what exactly will be evaluated and which indicators follow from that. Second is to explore the availability of data on the subject, followed by the selection of the most appropriate methodology. Based on the evaluation methodology, the fourth step is to determine the data requirements, followed by collection of the data and the sixth and final step is analysis of the data leading to a final report [2]. This sequence will be followed and then analysed for Social Impact Bonds to identify whether an evaluation tool can be developed to evaluate social impact and if so, whether this can be applied for Social Impact Bonds as well. But first, the concept of Social Impact Bonds should be introduced.

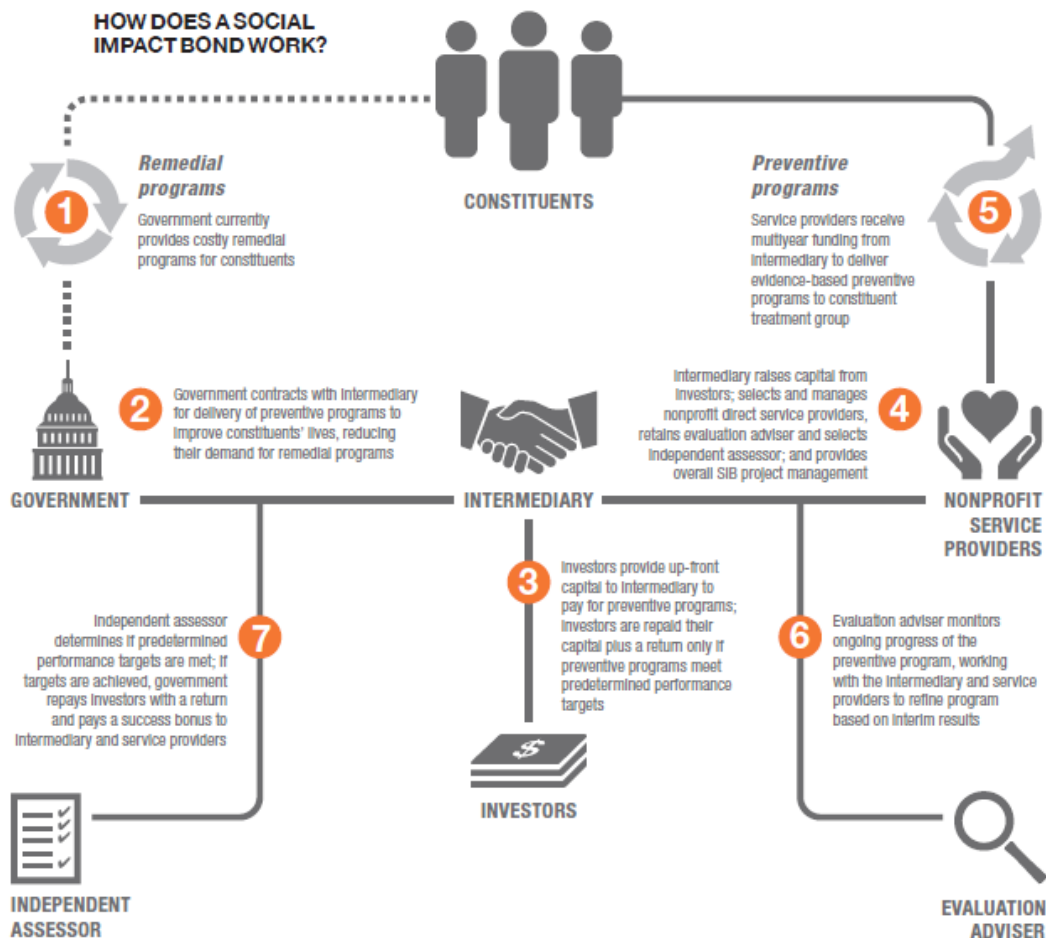
1.1. Introducing Social Impact Bonds

Before introducing the mechanics of SIBs further, it is important to define what they are. As the name might cause confusion (they are not bonds as the financial sector knows them), a clear definition is

needed. For this the definition as stated by Gustafsson et al (2015) is used.

a Social Impact Bond is a model in which private investors put up capital to fund a social intervention and governments repay the investors only if an agreed-upon outcome is achieved [33, p.4]

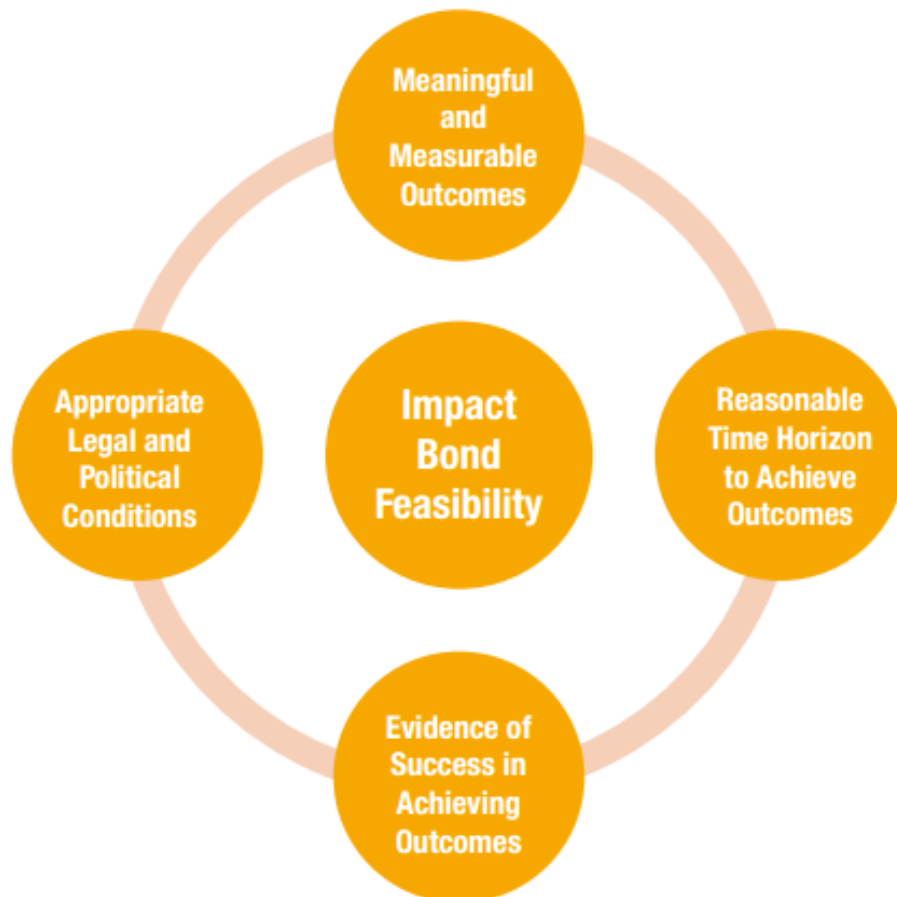
The topic of SIBs was first introduced in 2009 [71] and raised both critics and enthusiasm. When first launched, SIBs were positioned as an alternative for funding social service providers based on outcomes rather than outputs [71]. In other words, results of a service will not be determined based on the amount of people who made use of the service (outputs), but based on the results that the service users have reached (outcomes). To do so a structure was developed that is presented in figure 1.1.



The general structure of a Social Impact Bond as proposed by McKinsey in 2012 [71, p.14]

The structure of SIBs identifies five main stakeholders. Constituents are the populations that receive a service and can be identified as the population in need [33], examples are unemployed youth or re-offending criminals. The nonprofit service providers (henceforth service providers), will provide services to the population in need to support them in resolving their need such as giving them the tools to find jobs to overcome unemployment. The capital for those service users is provided by the investors. Fourth is the government, however this is now defined as the outcome funder as it not in every case needs to be a government party [33]. The outcome funder will agree on pre-determined outcomes that are aimed to be achieved for the service user and will pay investors if those outcomes are achieved. Finally the last main stakeholder is the intermediary, which is focusing on bringing all the other stakeholders together, managing their relationships and/or managing the performance of the service providers.

As the topic of SIBs is relatively new, the result is that there is also little extensive research known on SIBs. In 2015, the Brookings Institute produced the first comprehensive overview of all SIBs until now [33], which identifies 4 main criteria that are necessary for SIBs to become successful as presented in figure 1.2. The criteria are the definition of **meaningful and measurable outcomes** that should minimally be indicative of outcomes that are predictive of the life trajectory of an individual. A **reasonable time horizon to achieve outcomes** which is based on evidence from previous evaluations that the specified outcomes will occur within this time frame. The service providers should be able to show evaluations that show **evidence of success in achieving outcomes** from closely mirrored services. Finally, appropriate legal and political conditions are needed to ensure commitment and support from all stakeholders and to enable outcome funders to pay for outcomes beyond the year in which a SIB is launched (which is especially relevant for government budgets) [33].

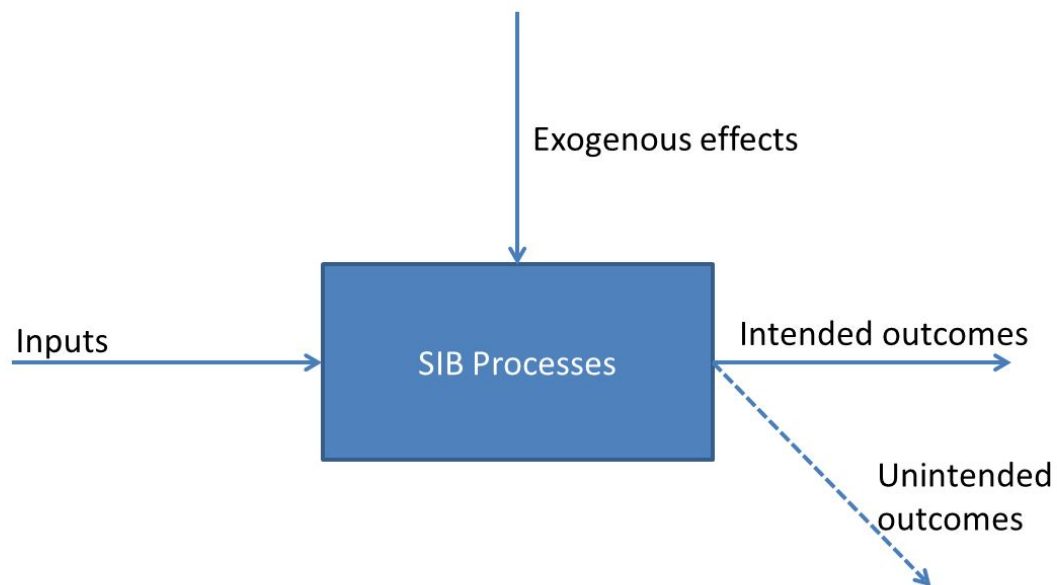


Four criteria to make SIBs successful [33, p.5]

1.1.1. SIB System

When SIBs are looked at from a system perspective, one might expect a system such as displayed in figure 1.3. This involves the actors within a SIB, which will be discussed in the actor analysis, but also the way the system interacts with the outside world and the different exogenous and endogenous effects.

The core of the system is obviously the box of the SIB processes. The processes can be considered as the intervention of the SIBs, the evaluation process, the interactions amongst stakeholders but also with participants. Participants for a defined for this thesis as the service users of the services that are provided by the service provider of the SIB. All processes that are directly involved with executing the SIB. The SIB then interacts with other parts of the system. First of all, there are two main inputs that will set up the SIB. The first input will be the resources that are needed before the SIB can start. This will be primarily the investment that is done for the SIB. The second input is the participants that will



A simple example of the system of a SIB

be involved in the SIB, all having certain characteristics in common based on the goal of the aim of the SIB (such as being unemployed, currently residing in prison, etc.). In between the start of the SIB and the realisation of outcomes, a SIB will be influenced by exogenous effects such as big environmental changes. This is important as those effect the eventual outcomes, but won't show the success of the intervention at all. Endogenous effects on the other hand are effects that are caused from within the SIB, when someone finds a job because of the intervention of SIBs. An evaluation tool will try to test causality in order to find out whether the outcomes that were identified, were actually a result of the SIB or from outside effects.

Finally, the outcomes are the effects that the SIB has had on the inputs. Financially, this will follow from the valuation of certain outcomes. More interesting are the effects on the participants, the actual outcomes such as going from unemployed to employed, from being lonely to being more socially active. The outcomes can be split in intended outcomes, effects that were intentionally and as designed by the SIB, and unintended outcomes, effects that the SIB has had but weren't expected. Unintended outcomes can be positive and negative. An example of a positive unintended outcome can be that when a SIB intends to improve employment and helps participants find a job, that in succeeding, the participant has a more healthy lifestyle (and potentially uses the health care system less). But unintended outcomes can also be negative. An example of a negative outcome could be that when rehabilitating ex-offenders and having them come to education or work placements on a daily basis, this is a positive direct effect. But when that ex-offender has a wife with a child who is used to having him available throughout the day, while making money off of drug dealing, that she will actually see it as a step back, making less money while having less time as well. This effect on her is an unintended outcome in itself, but can potentially also negatively influence the actual participant of the programme [57].

While the inputs, SIB processes (the intervention) and the intended outcomes are part of the expected system of SIBs and accounted for in its setup, especially the unintended outcomes and the exogenous effects are very interesting for the analysis of SIBs in general, but also for evaluating SIBs. Therefore they will be discussed separately.

Unintended outcomes

Outcomes that are the result of the SIB, but were not established as intentionally, can be seen as unintended outcomes. They can be either positive or negative, but significantly influence the outcomes of the effect. Both should be taken into account, but the difficulty is that they are often unintended, because they are not foreseen or too far out of the regular scope of a project. Therefore, including it into

evaluation becomes more complex as well. Therefore, each SIB should be evaluated based on a broader scope of well-being including qualitative evaluation in order to assess on which areas an intervention has made a difference, whether it is positive or negative, to deduct the potential unintended consequences.

Exogenous effects

Potentially more complex are the exogenous effects on a SIB or on any social intervention. Clear examples of exogenous examples could be changes of the educational system or a big financial crisis that will change the outcomes significantly. When in a financial crisis, unemployment rises, it will become harder to find jobs for participants of the service. This will influence the outcomes based on employment indicators. While the intervention itself might perform at the same level, they will achieve lower results. There are methodologies that will try to eliminate those exogenous effects by using control groups, such as randomised control trials, where the control group will be affected similarly by the exogenous groups. There are some ethical considerations in selecting control groups though, in cases where it might be the case that an intervention sets out to help all people in a certain population. In some cases exogenous effects can be measured by using available statistics on indicators as baseline data and correct for changes in those statistics (for instance a national increase in unemployment). The complexity of exogenous effects in social interventions is that people might participate in an intervention, but then engages in all other aspects of life on a day to day basis. This makes it close to impossible to rule out exogenous effects [54], let alone which exogenous effects have been taken place. What is important is that these exogenous effects have been thought through to some extent and that thinking is part of the design of a SIB. One should not try to eliminate exogenous effects completely though, as that is impossible to do.

There is also another reason to eliminate exogenous effects as much as possible, which is mentioned less often. What if the SIB fails is a question that can be hard for involved stakeholders on itself, but what if a SIB fails due to exogenous effects. For instance, an economic crisis hits and suddenly the targets of getting young people into jobs don't seem all that realistic any more. Even more reason why it should be tried to eliminate exogenous effects as much as possible from the evaluation of the SIB, but as also stated, it can't be done completely. This also means that it could still happen that SIBs fail due to exogenous effects. Establishing this fact also has an effect for the governance of the SIB. Who carries responsibility in those cases? Should it be government, as they are responsible for the general economic stability of a country? Most likely this should be the investors. They are also used to dealing with those situations when investing in regular investment opportunities. Those can fail due to changes in society as well, which is included in the risk analysis investors will (or should) perform on an investment. The risk of exogenous effects should therefore be limited, but also taken into account when investing into a SIB.

Causality of SIB

When the causality of a SIB is assessed, all the system interactions should be included to establish whether the inputs have indeed led to the intended outcomes and how well that has been done. In order to do so, one would look at all aspects of the system before causality can be completely understood. However, as discussed it is difficult to assess both the exogenous effects and the unintended outcomes. Nonetheless, it should be tried to control for the intervention in a SIB to see which effects it has and whether a SIB actually causes positive outcomes. To do this, all aspects of the system should be monitored and evaluated as thorough as possible, to get the answer to the question of causality as clear as possible.

1.1.2. Relevance

Currently the world experiences more inequality than ever before with the richest 1% owning almost half of the world's capital [34] and that gap has only been increasing, maybe surprisingly this widening of the gap continued even after the financial crisis. The current economy is focused on ever-continuing growth and that means that money within the financial sector will be always seeking to create financial returns to create capital growth. With the current wealth distribution this means that the gap will become increasingly wide, while too much inequality actually has a negative effect on an economy [80]. Since there is no shortage of capital for the World as a whole, this also shows that capital should be redistributed or more focused on increasing human capital. By looking at more possibilities to use capital to increase the real economy rather than the financial economy and use the excess of financial

capital to increase human capital through the concept of capacities, which is explained in the next chapter, this can make a big difference for society. The previously explained concept of SIBs is such an instrument where private financial capital is used with the aim to improve social challenges instead of just the growth of financial capital. Scientifically, the research is relevant as well. Currently there is hardly any research available on SIBs, while it is unclear how the concept compares to existing theories. The result is that there is a lack of critical thinking concerned with the development of SIBs. An aspect in which valid evaluation will play a crucial role to develop and improve SIBs where possible.

As a product, SIBs, when developed properly, also could have the opportunity of being relevant to broader topics. As discussed it can provide an option to capital to deliver more societal impact and have a positive effect on the real economy. It could allow the financing of a variety of topics, splitting the outcomes and the investment. Focusing on preventative solutions to social issues can make sure that innovation and risk taking is much more possible in this sector than previously. Therefore as a product it is innovating the world of social impact, but also that of investors. When relating it to more generic topics, a larger development can be found for instance within the European Union on the topic of responsible innovation. Within the Horizon 2020 programme, the strand of responsible research and innovation has been launched by the EU, enabling research and innovation that *implies that societal actors (researchers, citizens, policy makers, business, third sector organisations, etc.) work together during the whole research and innovation process in order to better align both the process and its outcomes with the values, needs and expectations of society* [22]. SIBs can play a significant role in this development, fitting all aspects of the goal of responsible innovation by bringing all societal actors together in one new process to achieve more social impact together.

The topic of Social Impact Bonds also brings an unique layer to the development of an evaluation tool. Obviously, there are several examples of evaluation tools on the topic of social impact, such as Social Return on Investment or Social Accounting and Auditing [99] or on the well-being of people [5, 87]. However, Social Impact Bonds adds extra complexity to the use of the different existing evaluation tools. First of all, interventions are not stable within SIBs. Different SIBs will have completely different interventions, targeting different social issues. This means that each evaluation potentially is slightly different. Secondly, SIBs consist of several stakeholders who have to negotiate on the arrangements of a SIB. While negotiation is an important part of modern evaluation [41], negotiation on each intervention that should be evaluated increases the complexity of designing a generic evaluation tool for all SIBs. Finally, evaluating SIBs runs a risk of bias due to the incentives that are accompanied with the different stakeholders. As financial payments are based on the outcomes of the evaluation, it is even more important that this evaluation is objectively as stakeholders have incentives towards a positive bias. This combination of factors makes evaluation of SIBs more complex than regular impact evaluation, but also shows the need for a well-designed and generic evaluation tool for these projects.

The structure of SIBs enforce a form of evaluation, as outcome payments are based on achieved outcomes and therefore have to be measured and evaluated in order to enable payments. Necessity doesn't however mean that there is no room for improvement or trade-offs that take place in this evaluation. Several items can be identified that an appropriate evaluation has the potential to improve upon. All stakeholders can benefit from well executed evaluation, but especially the commissioner and the service provider will benefit in the short term. Commissioners will gain a better understanding of their social problems and how to solve them, while service providers can improve their services while at the same time delivering proof that their services are successful (at least, that is the goal). When this leads to improved social interventions, they will benefit from the outcomes. Finally the investors might not directly benefit from improved evaluation, but they will require thorough processes to evaluate their results and make sure that they receive the financial return that they are actually entitled to. Even though evaluation currently already takes place, there are areas where it could clearly improve.

Evaluation has expanded to both outputs and outcomes, but does not fully include the perspective of participants on the effects of the intervention

SIBs bring an important next steps in the evaluation of social interventions. Generally, these are subsidised and funded based on their input/outputs, they get a certain amount of subsidy to run a programme and as long as they see X people within the set time frame, this method of evaluation doesn't say anything about the effects of the intervention though. Within SIBs, success is measured and payments are done based on the outcomes of the intervention, rather than the outputs. So not how many people have used the intervention, but for how many people has the intervention worked. This is an important step forward, but it is not enough and misses one crucial point. What are the actual

effects of the intervention on the individuals that participate. While finding a job or not engaging in criminal activity can, for good reasons, be seen as a good thing for society and are automatically also treated as positive for the individuals. However, there are circumstances where the individual might not perceive having a job as a positive over being unemployed. These perceptions are important to include, as well as their individual improvement in well-being. This aspect is not taken into account in most evaluations until now.

Evaluation is done primarily quantitatively but lacks qualitative assessment of the processes

Most indicators of SIBs are measured quantitatively, this is clearly necessary when you want to pay money based on a number of outcomes. The result can be that stakeholders will actively look for those numbers, but look past the qualitative aspects of evaluation that are just as important. Just as important as reaching successful outcomes for participants, is finding out what has led to that success and if or how those results can be improved. Evaluating processes is best done by qualitative rather than quantitative evaluation [2]. This qualitative evaluation should be implemented to look at the service providers and how interventions are delivered. Including throughputs and not just measuring inputs/outputs in the monitoring phase.

There are incentives for several parties to influence the evaluation process as they benefit from successful outcomes, evaluation should consider those incentives

All the main stakeholders have their own reasons to participate in a SIB, but those reasons might also influence the evaluation. For instance, the service provider benefits from successful outcomes (non-financially, but in some cases also financially), but in many cases the burden of proof also lies with the investors. This creates an incentive for the service provider to positively influence the outcomes, especially when the outcomes are subjective. These incentives are inherently included in a SIB with payments based on outcomes, but are not frequently evaluated. The roles between stakeholders and the effects of that on the processes of the SIB are currently not taken into account in evaluation, but is crucial when evaluating the system of Social Impact Bonds.

1.2. Research objective and questions

The above shows the necessity of an appropriate evaluation tool for SIBs. There are clear risks for the validity of the evaluations that are currently being performed, as well as in the system of SIBs as a whole. Poor evaluation will have an effect on the lessons that can be learned from earlier SIBs and reduce the trust of outside viewers in the results posed by SIBs. It is therefore critical that the validity of evaluating SIBs is looked at and an evaluation tool is developed that improves the validity of evaluation as well as taking all parts of the system into account. It is also clear that there currently is no such tool directly available. Therefore the objective of this research is:

To design and test an evaluation tool that evaluates Social Impact Bonds

To satisfy that objective, first a main research question should be established that will be answered throughout this thesis. Also, a set of subquestions will be defined for the different steps in the research and to support the main research question and the research objective.

Main research question: What are the design characteristics of a tested evaluation tool for Social Impact Bonds?

This question cannot be answered directly though. First it should be identified how an evaluation tool should look like and what criteria there are to its development. Therefore the first subquestion is:

1. Which set of criteria should be met to develop a generic evaluation tool for Social Impact Bonds?

If the requirements for an evaluation tool are known, the tool should be designed based on the established criteria. It should be answered which characteristics such a tool should have and when and how it should be used.

2. What are the characteristics of the designed evaluation tool for Social Impact Bonds?

As with the regular design cycle, the tool will be tested on the case of Social Impact Bonds to see whether the design is still applicable. This will be done by testing the tool with current stakeholders of Social Impact Bonds and get their view on the designed tool as well as doing a case study on Social Impact Bonds in their current form to explore to what extent the designed tool would be applicable to the current SIBs.

3. What are the results of testing the evaluation tool on Social Impact Bonds?

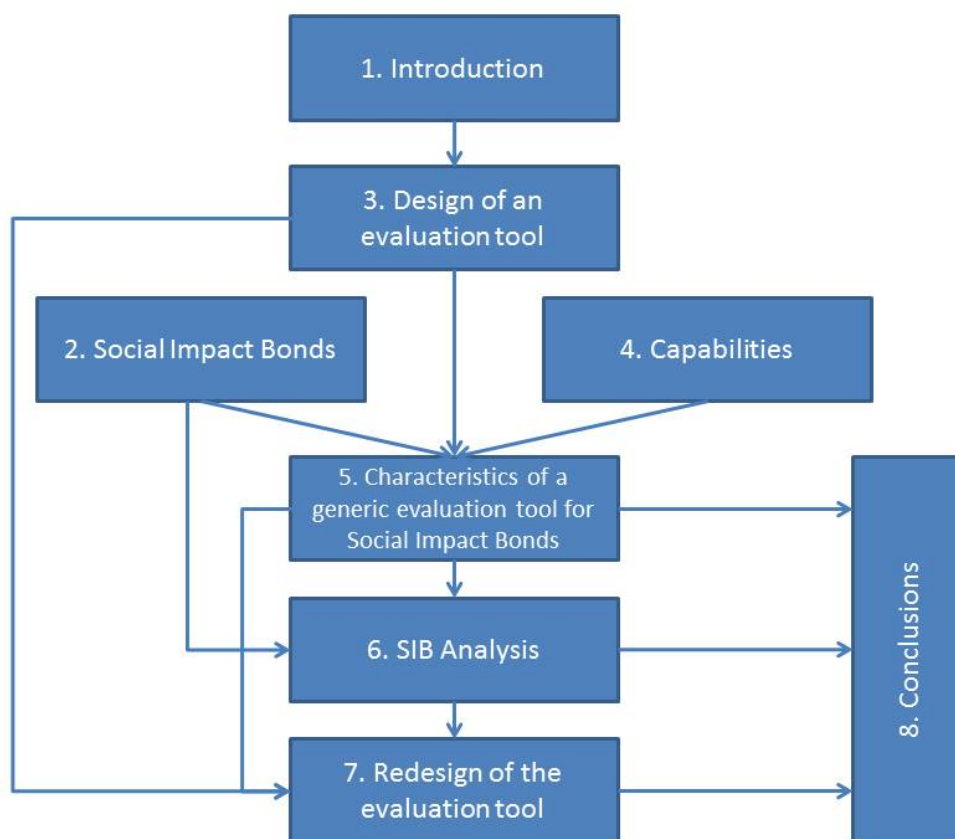
Finally, when the design is tested based on the case study of SIBs, it should be analysed which conclusions should be drawn and if and to what extent the evaluation tool should be redesigned.

4. How should a generic evaluation tool be redesigned based on the case study on Social Impact Bonds?

1.3. Structure of the thesis

This thesis will gradually answer the research questions that are set out in the previous section. For the rest of the thesis, first it will be investigated what criteria should be met to develop an evaluation tool for social interventions. Then a better look is taken at the SIBs and its actors, as well as at how measuring well-being can take place, for which the Capabilities Approach as initially developed by Amartya Sen and Martha Nussbaum will be analysed. Based on that input, an initial evaluation tool is designed. This will develop the characteristics of the tool and some generic aspects are directly designed. This design will be tested in two ways. First the tool is discussed with currently active stakeholders of SIBs to assess what their view is on the designed tool and potentially identify improvements to the tool. Secondly, the currently active SIBs are analysed as a case study to look into the existing SIB's systems, evaluation and stakeholders, and test whether the tool would be applicable for the variety of existing SIBs that have been launched so far. Finally, a redesign of the evaluation tool is proposed based on the tests and conclusions will be drawn upon the thesis to set out how the tool could be executed in practice and should be developed further.

In figure 1.4, the structure of the thesis is visualised and will also show how chapters are related to one another. The introduction will kick-off the thesis, but chapters 2 to 4 will all inform the design of the evaluation tool. The chapter about SIBs will also directly inform the analysis of SIBs that has been done in chapter 6. The initial design of the evaluation tool has also informed both the analysis of the SIBs and the interviews that have been held, but is obviously used as the input for the redesign as the new design will be based upon the initial version. The criteria for the tool, as developed in chapter 2 will also be used for the final design of the tool in chapter 7. Finally, both the design and the redesign and the analysis will form the basis on which the conclusions are written.



A flowdiagram of this thesis

Social Impact Bonds

With the initial introduction of SIBs in 2009 [71], it is a fairly new product. As can be seen by the quote of Toby Eccles, one of the originators of SIBs, it is also a product that was initiated from a practical perspective rather than being based on grounded theory or initial research: "I always find it interesting when people do research on SIBs as there has gone no theoretical thought in developing the model. It was a practical solution to a practical problem; we were trying to solve real-life issues". The result of this is that there is also very limited research available on the topic of SIBs, also partly due to the fact that there was some time needed for the topic to get known in the academic world and there currently only is a limited number of SIBs available. In this chapter, a more in-depth introduction to Social Impact Bonds will be given. Looking in more detail at the different actors within a SIB and their roles and responsibilities. Also, the underlying trends that have driven the upcoming of SIBs and potentially affect further growth of the concept are examined.

A Social Impact Bond is the name of a multi-stakeholder partnership between parties who will be called investors, social service providers, commissioners and potentially intermediaries and independent assessors. In general the structure will be in line with the presentation in figure 1.1. In more detail the role of each stakeholder will be discussed separately, but the seven steps provided by McKinsey & Company give a good first overview of the general process as it was first proposed in 2012. Usually governments provide remedial programmes for their constituents. Often the government pre-finances those services in order to be able to run. In a SIB, the same government will approach an intermediary or a social service provider for a new type of contract. In figure 1.1 this will be done through an intermediary. The intermediary raises capital from investors to provide the service provider with working capital, ideally on a multi-year basis. The results of the service provider will be analysed on an ongoing basis and in line with the SIB contract the government will pay the investors based on those results. The repayment is directly linked to the success of the service provider, therefore also the return on investors will be directly related to the success of the service provider. The risk of the programme is now also in the hands of the investors instead of the government, because if the service provider does not reach the agreed results or outcomes, the government's payment to investor might end up being lower than the initial pre-financing.

Throughout the world there are now 47 SIBs launched with investments totaling at least €95million¹, potential outcomes could get as high as €40 million. Although this is a small proportion of the whole impact investing sector and an even smaller proportion of the total money circulating the world on a daily basis, it is an increasing amount of money.

2.1. Actor analysis

As a partnership, the different actors in the system are crucial for the success of SIBs. Therefore, it is important to understand the different stakeholders, why they would participate in a SIB and what

¹This is the total investment for the 24 SIBs where the investment is known. Also SIB's have been launched in countries with different currencies. For the purpose of this thesis, a fixed exchange rate is used to convert to monetised values of SIBs to Euros, the exchange rates are GBP-EUR: 1.40, DOL-EUR: 0.89

their role is. For instance, when there are incentives for an organisation to behave in a certain way, it is important to know that upfront and handle that in the way agreements are structured.

An overview of the stakeholders can be seen in figure 1.1. There are four main actors involved: the commissioner, service providers, service users and investors. Without those four actors, a SIB couldn't operate. Then there are two other actors that are less vital, but can still play an important role and have so far been active in a proportion of the existing SIBs: intermediaries and evaluators. In this section, all those actors will be discussed to clarify their roles.

2.1.1. The commissioner

The first stakeholder to be discussed is the commissioner. This specific stakeholder is called the commissioner rather than government, as it not necessarily has to be the government, but could very well come from a variety of sectors. The main purpose of the commissioner is that it commissions the outcomes. That is also why it is such a central stakeholder, because it is a vital part of the SIB. As with each other regular business, services or products need to be sold in order to create income and keep the organisation going. For a lot of social services, it is more complex to create income as an organisation if you're helping ex-offenders to reintegrate into society as with the Peterborough example. The people on the receiving end of the service are not able to pay for the service, but they are not the only beneficiaries of such an example. In this case, society will see potential benefits as well such as reduced crime rates, lowered health care costs and lowered unemployment benefits and on that basis, the government is used to paying for those services. It is harder to measure those results but also harder to pinpoint the exact person or group of person that are receiving those benefits. That is where the commissioner comes in. The commissioner is a party that recognizes the savings as partly or fully being beneficial to them. This could be the government who will see a reduction in unemployment, but also a health care provider who has more healthy clients and therefore will see a decline in claims.

Until now, the commissioner has always been a governmental institution. The way they determine the money that they are commissioning, there have been broadly two options. The first is a straightforward link to the savings that are realized for them, which can be realized with unemployment programmes for instance. In this case, every Pound/Euro/Dollar that is saved will afterwards be commissioned to the other parties. The second option is that of a tariff structure. In that case, the commissioner will beforehand value the results of certain outcomes and commission that amount for each successful outcome that is realized, often up to a cap. This structure can be used in cases where the savings are somewhat unclear, spread over several departments or too far a way to be directly linked to short term savings. An example of the latter is an education programme where they will focus on keeping children that are in danger of becoming NEET (Not in Education, Employment or Training) on school longer and improve their school performance. If successful it is evidenced that having young people finish their diploma's has significant benefits to society, but they are realised in the long term. To be able to still commission those services, the commissioner sets a tariff on the value of outcomes as improving school attendance or finishing their diploma. For each successful result on a specific outcome, the service provider can now claim the outcome value of that. For instance, if the government sets a tariff of €200 on a rise in school attendance and the service improves the school attendance for 20 children of the target group, the commissioner will pay out €4000 for that. Ultimately, the valuation of these outcomes should be part of the mechanism where in a way, a commissioner is willing to buy services/outcomes at a certain price and investors and/or service providers must be willing to sell at that price. The determination of which outcomes to target and which indicators to use should be more open to discussion between the different actors though.

With a SIB the commissioner will be able to pay afterwards instead of beforehand as they do with commissioning regular subsidy's. This might create some uncertainty in their budgets though as it is not clear upfront how much they will be paying out for the programme. While the shift of risks away from the commissioner (and therefore often the government) is often seen as one of the major advantages, there are also some other aspects to it. SIBs have some impact on the budgets as there is now paid later but costs can also directly be offset by the savings, this also creates opportunities for scaling successful programmes more easily as there is not necessarily a budget increase because of an increase in scale. On the other hand, a SIB can also be much more costly than commissioning directly. The involvement of extra stakeholders creates extra costs relatively to directly working with service providers. These costs can either be accepted by a commissioner or can be included in the programme costs and then be shifted to the money that the service provider or investors will receive.

Commissioners have an interesting role compared to the other actors in a SIB. On one hand there is a clear incentive for them to identify successful interventions and scale them up. A government can be expected to create and grow social value, which is in line with the goals of a SIB. On the other hand there are some risks involved in working with extra organisations, as costs go up and it could be perceived as allowing private investors to make money off of the government. Therefore, it is important that they have independent evaluation to scrutinise the SIB, but also that they are able to set prices independently from the market, to avoid potential conflicts of interests. When using more commercial commissioners, it could be argued that they might have an incentive to downplay the outcomes of a SIB in order to reduce the amount of outcome payments for the same results. A commissioner also might want to select participants in a way that the largest social impact can be realised, rather than just creating successes for easier targets. Commissioners are usually governments, which will have the role of adding value to society as efficiently as possible, but is also expected to be very risk averse. Without SIBs they usually don't know exactly how much value they add and by giving out subsidies at the start of an intervention they will also fully bear the risk of a failure of the intervention. SIBs can bridge that gap to the desired situation by shifting the risk to investors and having a more direct measurement system based on indicators, to see what value is being added. A SIB is more costly due to the involvement of different parties though and therefore it needs to be ensured that their funds are still being used as efficiently as possible.

2.1.2. The service provider

The second important actor in a SIB is the service provider. Although the commissioner was presented as the central stakeholder in a SIB, in realising social impact, the service provider is obviously crucial. The service provider is an organisation, or even individual, that will provide services to a specific target group. For example, providing guidance to unemployed or organisations who meet released prisoners at the gate, training organisations, etc. The service organisations are responsible for delivering the social impact itself, for improving people's lives and indirectly for realising savings for the commissioner. Therefore the service provider usually will be assessed for its track record, organisational strength and potential as if the investors would do a regular investment. They often face tough challenges as they are trying to address social issues that are difficult to solve, otherwise we would not have needed service providers, or subsidies dedicated to improve society at all.

There are two options to improve the outcomes of a service providers from the perspective of a SIB and that is by scaling the service (so that it might take up more service users) or by supporting the service provider to increase the proportion of people that they achieve outcomes with. Where SIBs with a duration of one year will make not a whole lot of difference to the service provider, longer term contracts might. When the service provider can be guaranteed funding for more years it can focus on doing what it intends, providing services, instead of going through the process of looking for funding every year. On the other hand, they are potentially losing some of its budget towards other fees and organisations. Also, they will be much more scrutinized by investors both before arranging funding and during the SIB. Finally, they will be more accountable for their results as the results are now monitored more closely then before and in a case of bad results, their chances of receiving future are also reduced significantly. In general this will be seen as a bad thing for the bad service providers, but if done correctly it could also provide more scale and funding to the successful ones and stop funding for unsuccessful ones.

For a SIB, a budget will be determined for the costs of the organisation to be able to deliver the services. A business case will be developed in which the developed budget will be offset against the savings or the set tariff by the commissioner to assess if it will be possible to run a service that has the potential of returning more money than the initial investment. If the SIB is launched, the service provider will be provided with working capital in line with the budget to be able to provide the services. Often this is done in tranches to monitor the progress and steer in the budget that is actually needed.

While the role of service provider is crucial, it is also includes incentives to bias outcomes. Where a service provider in the subsidy system was just challenged on the number of people that participated, in a SIB it is based on the number of outcomes they realise. This might lead towards the choice of targeting the easier people, for instance out of the unemployment population the ones that are deemed most likely to find a job again, to generate more successful outcomes. This would however skew the results as the baseline group might take a much broader population. While this incentive might be aligned with that of investors, as it will generate additional outcome payments, it is certainly not aligned with that of

commissioners as it risks realising lower social impact for more money.

2.1.3. The service user

Service users are not directly seen as the central stakeholders, because they don't necessarily play a role in the partnership that a SIB is. However, in the end the service user is the most important element of a SIB as that is where the outcomes should be tried to achieve. If the aim of the SIB is to improve social outcomes or build capabilities (as researched in this thesis), this also means that the service users are essential in that process. Those outcomes and therefore the type of service user that is involved can vary over different SIBs. For instance, with the SIB in Rotterdam it will be young people aged 18-27 while with a SIB around end of life care, it is people who are near the end of their life and therefore much more likely to be older. The same will go for social status or gender. Nowhere is there a set definition of the type of service users that can participate in a SIB.

Usually service users will participate in programmes that are provided by the service providers. As mentioned, the programme they participate in will be very dependent on the type of SIB and on their needs. This can range from attending a programme during school that pairs teens of 14-15 with toddlers and teaches them responsibility in weekly sessions where they have to take care of that toddler, to meeting prisoners at the gate when they are released from prison and helping them develop a plan towards reintegrating in society and stopping them from re-offending. Service users can enter programmes by applying for it themselves or by being assigned to it (i.e. by governments or schools). There have been no cases so far where service users were obligated to participate in the programme, even in cases where a service user was assigned as suited for the programme by a third party, they were still free to not participate.

The roles and responsibilities of a service user can differ strongly over the different types of SIBs. In some occasions, service users can apply for interventions which will obviously create more committed participants than when they are appointed to participate. In some countries/areas, unemployed people might be forced to participate in employment programmes, in which it might be seen more as the responsibility of the service provider or the local government that service users enter and remain in a SIB. The way in which service users enter a SIB can highly influence participation rates, but also determine outcomes. When a service user participates and gains an advantage when the service is successful, he/she might be more willing to cooperate. When it is just an obligation in order to remain eligible for unemployment benefits, the service user might be less interested in the positive outcomes, which reduces the likelihood of those outcomes being achieved. In general though, the goal of a SIB is that a service user sees an improvement in the services that are offered. As the service provider now has an incentive to realise outcomes, the service users also should expect more targeted services and therefore more improvement to their lives.

2.1.4. The investor

Investors are the fourth essential party to a SIB. Now the commissioners and service providers are explained, the investors will be the parties or individuals providing the funding for the SIB. There is no direct definition of investors here, as with the other parties there is not just one type of investor. In a SIB it is the organisation or individual who provides the funding for the SIB at the start of the project. If and how they will receive their money back can also differ depending on the specific project. In previous and existing SIBs investors have been banks (Triodos or ABN AMRO), investment funds who are dedicated to impact investing (investing with an aim for both financial return and social impact) or foundations who have set aside some of their funds for impact investing. Investing money in a SIB is usually done with the aim of making a return on investment while at the same time creating social impact. Usually potential returns on investment are in the range of 0-10% IRR which is significantly lower than what for instance general private equity funds aim for, but obviously lower than most foundations are used to when they are giving grants.

Financial organisations and investment funds are more and more asked to strive for a more socially oriented portfolio. Not in the least by the public opinion, but also amongst the wealthy individuals who are usually providing the money for investment funds a trend can be seen for more socially oriented investing or use of money. The giving pledge that was first started by Bill Gates and Warren Buffett has now increased to 131 billionaires that have signed up to give away more than 50% of their wealth over the course of their lifetime [55], which is an example of this trend. SIBs are a form of impact investing, which in itself fills the gap between regular investing with the aim of the highest financial

return possible or giving money away in order to *do good*. Therefore, the growth of SIBs or impact investing as a whole might be driven by a demand pull from the investor size. The result is that SIBs provide an opportunity to be in the middle of those two options with a modest potential return. It is however unclear how it compares to the two other options in terms of social impact, there is the goal of social impact which sets it apart from regular investing, but in terms of actual social impact there is no comparison standard available with for instance grant giving. Based on the available wealth data through Forbes, at least 50% of an estimated wealth of \$600 billion will be given away, which compares to 50% of the GDP of the Netherlands [30].

Even though, investors in SIBs are often socially motivated, they should be treated as mainly having interests in financial return. This has effect for their role in a SIB, as it then can be perceived to be optimising outcome payments and using inputs as efficiently as possible. The dynamic for the negotiation of evaluation/monitoring can therefore be interesting as the incentive to optimise outcome payments is directly aligned with the service provider, while minimising inputs actually goes directly against the service provider. This means that clear agreements should be made around both items to make sure that an investor cannot influence either one directly without consent of all stakeholders, but without limiting an investor too much, as pushing for more efficiency might be completely justified in cases and therefore have a positive effect on the SIB as a whole. While a SIB doesn't necessarily bridge a gap to a desired situation, it is part of opening the market for more investment opportunities in the space of impact investments. For more commercial investors it might bridge the gap from the desires of their clients, if applicable, to invest more in socially oriented opportunities, and get a return as well, while previously it had often been one of those two things.

2.1.5. The intermediary

The role of intermediary is one that is not compulsory for a SIB, but organisations such as Social Finance UK highly recommend it. The function of an intermediary can be split up in mainly two phases. In the engineering phase as you could call it, the phase where parties are brought together, agreements are being made or in some cases bids are being done for a tender, the intermediary will coordinate parties. Multi-stakeholder negotiations can become complex and the intermediary could play a role in streamlining that. Once the SIB is up and running, the intermediary can play a role in monitoring the progress and results of the service provider and supporting in the reporting area of that towards commissioners and investors as most service providers are not used to the expected level of reporting. With putting such a task on an organisation, there are obviously costs involved which will also make the SIB more expensive if an intermediary is involved. This could be justified when for instance these extra costs are offset by improved results by the service provider. To what extent that is actually realised is also yet to be proven.

Usually an intermediary has a fixed fee, either for setting up the SIB or for specific resources in the performance management role. This means that an intermediary can be an independent party leading the negotiations between stakeholders. However, usually payments are coming from the contract funded by investors and also used by the service providers. Therefore, the role of an intermediary might be different towards investors/service providers than to commissioners. Nonetheless, when intermediaries do not have financial incentives, they can be expected to act as a third party to make sure that the aims of the SIB are fully satisfied. In cases where intermediaries do have financial incentives, they can be expected to behave in a similar role as service providers within the SIB.

2.1.6. The evaluator

Several other parties could be involved in a SIB although they are never essential. In big deals there can be legal organisations involved to structure agreements or deals. In evaluating results there is also an independent assessor needed as described in figure 1.1. Although this is not a strict requirement for the SIB, it brings extra value in itself because the service provider, commissioner and investor(s) all have different incentives which are not necessarily aligned with each other. To make sure that results are interpreted independently, external assessors might be used to make sure the outcomes of the evaluations are valid and can't be biased by an individual stakeholder. Not all SIBs have been making use of an external stakeholder though. Also, the criteria for evaluating SIBs have shown that an independent evaluator can increase the validity of the evaluation.

2.1.7. Conclusions of the actor analysis

One of the issues with such different stakeholders is that of managing the relationships between all stakeholders. Investors might be very oriented on social impact, but could also be mainly focused on making a return on investment. Different incentives or objectives could cause for tensions within the group of stakeholders of a SIB, but also towards biased outcomes or cherry picking of participants. Mainstream investors that are making significant returns off social issues, funded by the government, might not go too well with the public opinion as well. Therefore it is very important that the objectives of each stakeholder is clearly stated and the different incentives are either aligned or become transparent. It emphasises also the importance of third party evaluation and potentially intermediaries to avoid bias and make sure that all stakeholders are involved in the progress and learn from the project, while at the same time negotiate with the other stakeholders.

To refine the 'clash' of stakeholders though, SIBs to date have also seen many firsts. The first SIB ever, the first SIB in a different country, the first SIB on a specific topic, etc. This means that the learning costs of launching SIBs like that have been relatively high as well, including the reputational risks of getting involved in such a public-private partnership for all stakeholders involved. It can be assumed that organisations, even when they are deemed to be very commercial or having certain incentives to skew the results of a SIB, are involved because they are intrinsically motivated to make SIBs work. This also involves collaborating with the other stakeholders and making it work together, instead of trying to influence each other for your own good. At no point does this mean that there are or have been no risks involved with the involvement of stakeholders and their perspectives, but in the early stages it is very reasonable to assume that the commissioners that have been involved are more likely to be innovative and forward thinking than their peers. The same goes for the investors, most likely they are more intrinsically motivated to realise social impact than the average investment fund or organisation. When SIBs will scale, these first movers will be followed by more regular organisations in the market place though. Making this point in time for SIBs even more important. As a system and as first movers, it is most important to define what is good and what not for SIBs. What is a SIB, how does it work and how do we evaluate success. When certain aspects have been defined properly, this reduces the risk of wild exploration of various forms of SIBs that have nothing to do with initial good motives or setup. At the point where stakeholders get involved with the topic of SIBs, purely because it is a good business opportunity (using the broadest interpretation of business), this will put a risk on the external perspective of SIBs and their stakeholders.

2.2. Trends that drive implementation of SIBs

Social Impact Bonds also fall into two bigger trends that expand further than just the work of these social initiatives that are being funded by SIBs. The topics of Impact Investing and Payment by Results have been discussed widely and are growing in scale. Therefore it is worth spending some time on those two topics.

2.2.1. Impact Investing

As mentioned earlier in this section, the fact that impact investing in itself has established itself as a global subject can be seen by the fact that the G8 Taskforce spent a complete meeting on the topic [26]. The Global Impact Investing Network (GIIN) defines impact investing as *Investments made into companies, organizations, and funds with the intention to generate social and environmental impact alongside a financial return. They can be made in both emerging and developed markets, and target a range of returns from below market to market rate, depending upon the circumstances* [62, p.11]. JP Morgan and GIIN take an annual survey amongst their members that are investing money qualified as impact investments and the 2015 survey shows that in 2014 alone, their group has committed investments of \$10.6 billion and intends to grow this to \$12.2 billion in 2015 which would be a 16% growth. Moreover, the respondents of their survey have a total of \$60 billion of assets under management, of which 35% is proprietary capital (or money owned by the respondents themselves) and 65% is managed on behalf of clients [62].

2.2.2. Payment by Results

Obviously Social Impact Bonds is not the first example of change happening in changing public spending. There are private initiatives trying to realise social impact (for instance with NGO's), but also

governments are looking at new ways to use their spending, for instance by making new types of agreements with service providers that are funded by the government. One of those types has a lot of similarities with SIBs while it also gained wide recognizing thanks to endorsements by Barack Obama, Payment by Results [44, 69]. With this scheme, governments of commissioners change their agreements on how to fund/subsidize projects. Instead of defining a budget and financing the projects upon the start, no pre-financing will take place. The government makes an agreement with the service provider that requested the subsidy on its results and if the results are actually realised, then the government will pay for only those results, hence Payment by Results. This is systematically very similar to that of Social Impact Bonds, but it differs in that there is no pre-financing of the project by the government or any other funder. This means that either the service provider itself needs have money to pay for its working capital, or needs to get a loan for the money from an external financier. In that case, the service provider will be responsible itself to make sure that there is enough working capital to provide the services and it will also carry the risk of not realising the results itself. There are both pros and cons to this method in comparison to the described system of a SIB and SIBs can indeed be seen as a form of Pay for Success [69].

2.2.3. Future?

History has shown that predicting the future is simply not possible. However, it appears clear that SIBs or similar social initiatives are here to stay and can be expected to grow in size in the foreseeable future. In the current times, where the richest of the world commit their money and new generations grow up who are not just focused on themselves but are willing to improve the world. This trend can continue for both sustainability themes and social activities. Enterprises, investments and therefore methods to make the best use of our resources therefore will keep growing and Social Impact Bonds are one example in that trend.

Design of an evaluation tool

Before the evaluation will be designed, the basic criteria for an evaluation tool should be set. In this chapter, literature and several specialists are counselled to explore the complexities of evaluating social projects and inform the design of the tool. Several factors will be looked at, starting with identifying the need for evaluation, establishing the basic criteria for an evaluation tool, and based on the steps as proposed by existing literature, the requirements for the several design steps are identified. This all will lead to a list of basic criteria that will be used to design the evaluation tool later in this thesis.

3.1. What is evaluation and why is it needed?

When the word evaluation is looked up in the Oxford Dictionary, the definition given is *The making of a judgement about the amount, number, or value of something* [27]. Therefore, to perform evaluation correctly, it should be established which amount, number of value will be judged, how it will be judged and by whom. When the concept of evaluation is applied to social impact, it is defined by the World Bank as *Impact evaluation seeks to determine whether the project has the desired effects on individuals, households and institutions and whether those effects are attributable to the project* [2]. In her paper, Adam separates monitoring from evaluation by defining monitoring as collecting and analysing data to see whether the implementation of a process is as designed and adjust where necessary, while evaluation is an objective assessment of a project to measure the relevance and achievement of its objectives. But both are done to improve the performance of a project and measure its results [2]. A similar distinction is made, by splitting it up in a summative function, which measures performance, and a formative function, which utilises the evaluation as learnings for future initiatives [41].

Social impact is something that can be achieved on a very broad scale. It can be reached by educating girls in Africa, or reducing the carbon footprint of large commercial corporates in New York, and everything in between. But the type of impact (where the first can sometimes be broadly defined as social impact, the latter is clearly more environmental impact), the scale of impact, and the beneficiaries are all slightly different. Evaluation is therefore needed for several reasons. First it is important to establish whether the projects can be improved in any form, then in some cases it is important to establish whether projects even have a net social impact or if they do, whether that social impact is worth the investment and finally to be see which projects perform better than others.

Evaluation of social impact can take place on levels of inputs and outputs (monitoring) or on outcomes and impact (impact evaluation) [2, p.6]. They will both serve different purposes of the evaluation process. While the impact evaluation takes evaluation an important step further from the evaluation processes originally used for social interventions, monitoring is just as necessary. Just measuring on inputs and outputs won't show the differences that are made. In practice, managing on outputs would mean looking at the number of times a social worker visited an unemployed young person within a certain time frame. Measuring on outcomes would look at the number of young people that have participated in a programme that actually ended up in employment. The latter comes much closer to actual social impact and therefore evaluation of any type of project that has the aim of realising social impact should be using impact evaluation aspects. This is in line with the summative function as described by Kuhlmann. The formative function of evaluation, to produce learnings from the processes are just as

important though and can be found more in the monitoring aspect. Whether the duration of a project is long enough to implement learning in the current project or are used for future projects, monitoring can provide insights into the inputs and outputs of the processes that can be used for potential improvements of the project. Therefore, a well-designed evaluation tool, should combine both monitoring and impact evaluation.

3.2. Criteria for the design of a tool to evaluate Social Impact Bonds

It is important to establish what criteria an evaluation tool should satisfy in order to appropriately measure social impact. This section will set out those criteria. Starting with the timing of the evaluation, establishing which requirements there are for the evaluator, followed by the implications for the validity of the tool. Finally, also the practical achievability will be looked at, as the costs and resources needed may bring up limitations to the actual design of the evaluation tool. In general, an evaluation tool is expected to test if the programme is appropriate, the target group is reached, what the impacts are, the targets are attained and if the implementation was efficient [41]. But it should also test whether the project can be improved, how the project compares to alternatives and how it affected the beneficiaries [2]. Especially the last part is important, not only because programme evaluations often overlook the interests of the citizens' groups [41], but also because social interventions often aim to improve the situation on an individual level, person by person, but are then evaluated on an aggregate level. Projects however should be evaluated on all levels, individual participants, project outcomes and the inter-stakeholder agreements [28].

When looking back on the four success criteria for a feasible SIB from figure 1.2, two are outside of the scope of evaluation, while evidence of success in achieving outcomes and to a lesser extent meaningful and measurable outcomes are directly related to the evaluation. While the measurable outcomes mainly inform the evaluation tool and therefore the indicators that lead to those outcomes should be constructed properly. Evidence of success however can only be found when the evaluation has been performed appropriately and the evaluation was valid.

Criteria on methodology and data collection should be formed in a way that the evaluation takes into account constraints on resources and the availability of information [41]. As social interventions can be very diverse both in scope and scale, a tool that is to elaborate might risk not being used for smaller scale projects, therefore the tool should likely be relatively easy on resources, both in human and financial capital. In an ideal world, the evaluation tool should use both control groups and baseline data, as well as mixing quantitative and qualitative techniques [2]. In reality, a generic evaluation tool that uses all those components, is probably not achievable. Especially live control groups can be costly and extremely difficult when interventions are performed on a large scale. When interventions are performed on a national level for instance, one would like to see the intervention to not exclude people to create a control group, while creating a control group from other geographical areas might not be valid or result in ethical dilemmas. Also causality in social interventions will be hard, if not impossible, to appropriately test due to the variety of exogenous factors [54]. It appears more sensible to create an evaluation tool that combines baseline data by doing before/after tests on the participants of the evaluated interventions.

The use of both quantitative and qualitative techniques can also phase into the two levels of measurement, outcomes and impact. While outcomes (i.e. employment levels) is something that can be measured appropriately in a quantitative manner, which will improve objectivity of the evaluation tool, impact is much harder to measure quantitatively as this will look at the perspectives of individuals. Therefore, impact should be measured qualitatively. The qualitative evaluation should however not just focus on the well-being of the individuals, but is also suited to look at the processes of the project and should therefore also look at the potential for learning, the formative function of the evaluation. Qualitative analysis will provide more true insights in the social aspects [54], while using hard objective criteria through quantitative analysis proves more honest evaluation without bias [90]. To enable comparisons between different projects or interventions, an evaluation tool should be generic to some extent. Social projects however are, as mentioned, very diverse, which makes generalisation difficult. Outcomes are measured quantitatively, but it is simply impossible to directly compare a reduction in recidivism with an improvement in school attendance. The most concrete way towards comparing those outcomes is by monetising outcomes. Monetisation of public value creation is a topic that has

been debated over and is hard to do accurately. When however, direct outcomes are measured by clear indicators, financial value can be attached to outcomes. The most straightforward example of that is (un)employment, where a reduction in unemployment also (in most countries) leads to a direct reduction in unemployment benefit payments, this reduction can therefore be monetised. Other social subjects are harder to monetise and the value one attaches can be subject to a lot of other factors than just the outcome. Improving the same outcome in different geographical areas or for different age groups, might lead to different valuations of that outcome. For the evaluation tool's objectivity it might be wise to separate the monetisation process from the evaluation process. Most social projects involve stakeholders with different incentives and perspectives, therefore the monetisation is subject to the relevant environment and should be set by the outcome payor who should set a target on their valuation of an outcome, because of the subjectivity of this valuation, it should not be directly part of the evaluation [28, 90]. The evaluation tool should test outcomes that have been monetised appropriately. Comparability of qualitative processes can take place by measuring the well-being of individuals through qualitative tools. Examples of measuring well-being on an individual level can be found in the application of the capabilities approach [5, 58].

3.2.1. When should the evaluation take place

When looking at the evaluation tool it is important to look at the timeframe in which outcomes and impact are realised [2]. The time it takes for actual outcomes to be realised can take a while, sometimes long after the actual project or intervention takes place. In those cases, establishing causality becomes even harder as the potential exogenous effects can influence the trajectory of individuals even more. The risks of these exogenous factors will either have an effect on the investors, when outcome payments are only made when the actual outcomes are realised, or on the outcome payer when payments are made at the end of the intervention [90]. An evaluation should however take the outcomes into account, even when they are long after the intervention, and evaluate the results, this reduces the practical achievability of the tool however, as there are several costs involved with the extended waiting period [28]. Therefore it may not be wise to evaluate after the outcomes have taken effect (or are expected to take effect), but rather evaluate directly before and after the project takes place. Indicators that are measured should in those cases be established so that they are predictors for the future outcomes and the rationale behind the choice/construction of those indicators should be transparent and taken into account when evaluating the project. In general, the evaluation should be included from the early stages of the project design to ensure that the before tests can be done appropriately and in a timely manner.

When monitoring, the evaluation obviously should take place during the intervention. As the function of monitoring is to provide learnings, monitoring should take place at moments when there are learnings to take in and ideally when there is enough time remaining in the project, to implement those learnings. Therefore, monitoring evaluation should take place at either a fixed time interval, or even better, in between fixed cycles of the projects. This monitoring should also be a constant topic of discussion amongst the involved stakeholders, to discuss learnings and improve the project methodology when possible [28, 54]. This leads to a constant negotiation on the implementation of the evaluation learnings [41]. For instance when a project works with different groups that start sequentially, in between the ending of the old group and the start of a new group would be a perfect time for monitoring inputs and outputs. When the projects does not appear to have those clear cut-offs, a fixed time interval should be taken that is reasonable for the type of project or intervention and should be completed consistently throughout the project. This time interval should not be too short (i.e. monitoring every week, while the project meets with participants only every month) nor should it be too long. The choice of the time interval will also influence the costs of measurement, as more evaluation moments will obviously drive up costs. Another big part of the monitoring is that the pretest for impact evaluation, should also be done for monitoring, to ensure that all parties are confident that the resources that are being put into the project, will be sufficient to realise the goals.

3.2.2. The evaluator

In the development of a generic evaluation tool, it is important that it is constructed in a way that it can be easily replicated and used. It remains important however to establish who will perform the evaluation, to ensure objectivity and rigour. Where, if the indicators are constructed appropriately, the quantitative evaluation ensures some level of objectivity, the qualitative evaluation has higher risk

of losing objectivity and being influenced by the evaluator. Therefore, criteria should be in place for who the evaluator should be for especially the qualitative evaluation. Two crucial preconditions can be identified. Firstly, the evaluators should have a high level of expertise, both methodologically and on the (social) topic at hand. Secondly, they should be independent from core decision makers, either politically or within the project [41]. This automatically leads to the criteria of a third party performing the qualitative analysis, which has the potential of increasing costs, but also increases trust in the findings and the learnings that might be taken from them. The selection of this evaluator should take place in the design phase of the project and it is important that all relevant stakeholders take part in selecting the evaluator [2]. As it is not feasible of involving every individual beneficiary in this project, the providers working with/for the beneficiaries should represent them in that process.

3.2.3. Validity of the evaluation tool

While evaluation ideally should take place with the use of a counterfactual or control group, to create the possibility of using methodologies such as experimental/quasi-experimental evaluations [2], or the use of randomised control trials [90]. This might not be feasible practically, while fully proving causality in such complex environments will be an issue anyway [54]. Before/after comparisons and qualitative analyses are also seen as very important evaluation methods [41]. Using before/after comparisons, predispositions and innate abilities of the population will be constant, increasing the validity of the evaluation [2]. It is however harder to measure effects of exogenous influences such as macroeconomic shocks. Evaluation results can be significantly strengthened when baseline data is collected prior to the implementation of a projects [2], re-emphasising the necessity to come to clear agreements on the evaluation of the project early in the design phase of a project. Using quantitative methods to analyse outcomes will increase validity, however the construction of indicators for the project should be looked at critically and should be treated carefully in instances where the actual outcomes might take effect after the project. The construction of the qualitative evaluation should also be treated carefully, as risks for validity are high. The development of a questionnaire can influence answers and should therefore be done with care to increase the validity of respondent's answers. In surveys on social issues, there is an increased risk for the validity as it is likely to touch upon topics such as income or drug use, that can be seen as sensitive. Such sensitive questions tend to lead to larger non-response rates as well as larger measurement errors [84]. It may trigger *socially desirable* answers from respondents that will skew the evaluation. To improve the validity of the responses, it is important that questionnaires will be self-administered rather than administered by an interviewer so that the interviewer will not see the respondent's answers [84]. Validity should also be increased by other parts of the design of the evaluation tool, the timing of the evaluation, who the evaluator is and using the appropriate evaluation methodologies.

Causality

Another important is that of internal validity, or cause and effect relationships. An appropriate evaluation tool should measure cause and effects relationships to see what difference the project actually makes. In order to do so, the evaluation tool should, on top of the aforementioned criteria and methodological requirements, establish causal models beforehand. Through these causal models, the effects can be tested, but it can also be established what other effects the intervention has that weren't expected (side effects). This will allow to potentially look at effects that are caused by the intervention that weren't desired, whether they are positive or negative. It is much harder however to evaluate whether there weren't any external effects that influenced the outcomes of the project. Participants within interventions constantly interact with the outside environment and it is therefore never possible to fully exclude external effects like you would in a true experiment. The method with the highest internal validity would be to use a quasi-experiment, by working with a control/comparison group. Working with a baseline would bring lower validity as external shocks (i.e. financial crisis, natural disaster), might influence results. Finally, qualitative evaluation can also provide deeper insights into the process and the effects on individuals, therefore giving more information about potential causality [54].

3.2.4. Practical achievability of the evaluation tool

Ultimately, each evaluation tool will face a trade-off between costs and rigour. The ideal evaluation tool, will obviously produce the most sounds evaluations, but can also be very costly. Costs that not all projects can easily bear, both financially as in resources that are needed to perform such an evaluation.

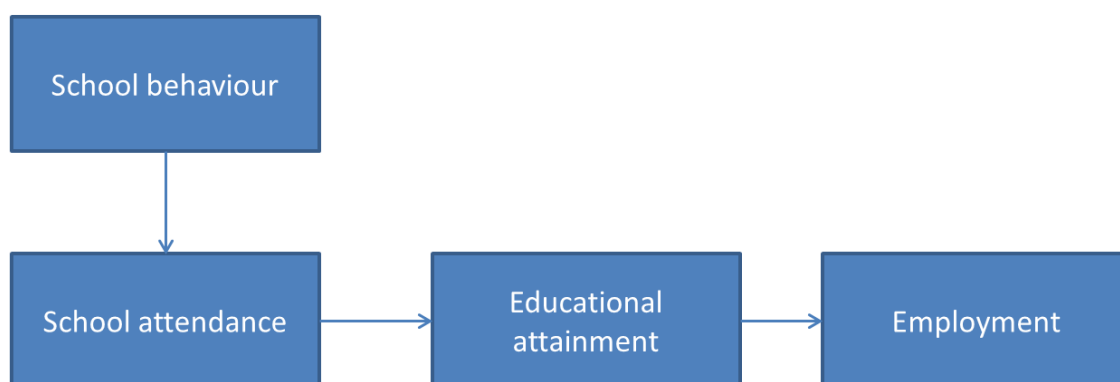
Therefore, choices have to be made around the different components of an evaluation tool. In general, expenditures for impact evaluations range from 0.25% to 2% of the total costs of a project, which means that on a project the size of a €1 million, still only €20K will be available for evaluation, which limits the possibilities. For projects like SIBs, where evaluation plays such an important role towards financial payments, it can be argued that evaluation and monitoring could take a higher proportion of the total costs in order to allow for appropriate evaluation. Therefore, to ensure that the evaluation can be carried out properly across projects of different scales, it will be aimed to develop an evaluation tool for which the total costs do not exceed an, arbitrary, cut-off of 4% of the total costs of small projects, for which the size of a small project is yet to be determined. When costs for evaluation become a larger proportion of the total costs, this will lead to a lesser focus on the actual execution of the project and endanger the achievement of potential outcomes.

3.3. Design methodology

The broad criteria from the previous section should lead to a more specific design of an evaluation tool that can be tested on a case study of Social Impact Bonds. In order to do this, the tool should be developed in a manner such that it meets the criteria as established previously, but can be managed properly with the reality of limited resources in many cases. The design will take steps to look into the specific evaluation criteria and/or indicators, the data availability and requirements, measurement methodology and finally data collection and analysis. The final step will include setting up criteria for testing and executing the tool in practice.

3.3.1. What to evaluate and which indicators to use

Selecting specific indicators for evaluation is hard without knowing the specifics of a project. Based on the established criteria for evaluation in the previous section, more specific requirements for the indicators can be set. Indicators should be relevant, selective, clear, specific, collectible and variable [2]. The c data will be discussed in the data availability and requirements section. To get a better sense of causality, the first step towards finding the specific indicators would be to develop a causal model. An example of a simplified causal model can be found in figure 3.1. In this diagram, it can be found that improving employment for participants through education is directly related by (secondary) educational attainment. To increase school attainment, school attendance and school behaviour can influence the results of young people both directly and indirectly. While this is a simplified example, it shows three indicators that can be used for this project, educational attainment, school attendance, and school behaviour. In practice, more sophisticated causal models should be designed that will take into account both exogenous and endogenous factors of a project so that they can all be analysed to evaluate all effects on the project [28].



An example of the causal model between education and employment

Using a causal model, a meaningful number of indicators can be selected that are relevant to the project. Where necessary, indicators should be defined more precisely in order to be completely clear. Using the above example, school attainment and attendance can be unambiguously measured as those

are quantifiable results. Behaviour is an indicator which should be defined further to be clear and unambiguous. This could be done, by scoring the behaviour on a specified range (i.e. 1-7 of 1-100), which then should be repeated on a regular basis by the same person. The development of the behavioural rating can then be clearly measured, but priority should be measured to indicators that are unambiguous and clearly measurable over indicators that are defined subjectively [90]. When the causal model is compared with the system diagram, the causal model focuses on the intended outcomes. When the constructed causal model, side effects that might be foreseen as resulting indirectly from the intervention should be included, to make sure that all intended outcomes are considered and not only primary objectives are used. Doing this properly will improve analysis and evaluation, as then intended outcomes can be truly separated from unintended outcomes. When unintended outcomes are found, the question *Could we have foreseen this outcome* should always be asked and be part of the learnings.

Once the criteria are selected and are constructed clearly, the next step is to make sure that they are comparable with other projects in order to make the evaluation tool not only generic but also providing a method to compare different projects. The most reasonable method is monetisation of the indicators. While monetisation is executed differently in tools like Social Return on Investment, but most social values are subjective and should be treated as such [43]. Therefore, the valuation of indicators should be just as much part of the negotiation as the evaluation itself and agreed upon by all stakeholders [41]. Involved stakeholders have different incentives, where governments might want to keep prices down to reduce their payments, investors want to use higher prices to increase their potential return. This negotiation is more likely to lead to a specific valuation that is appropriate for the project, but in the end it should be the outcome payer that determines the final valuation [90].

The indicators constructed from the casual model only measure outcomes, while it is established that also impact should be measured. Impact is seen as well-being of the beneficiaries, and should therefore be measured on an individual level. Measuring well-being is something that has been done before using the Capabilities Approach [5, 58], which is also the basis for the human development index [87]. The capabilities approach is very broad, but can be used to measure overall well-being, as well as changes on specific capabilities. The measurement of this can be generic for every project, but is separate from the outcome indicators. The existing measurement method uses a quantitative questionnaire, but to properly measure effects and get deeper insights in the process, they should also contain qualitative questions [54].

Finally, the monitoring should take place to implement learnings. Monitoring is based on inputs and outputs of a project, but also should include monitoring on outcomes when the project time lines allow for it. Inputs and outputs can be similar for each project, inputs being the resources that are put into the project, both in people (FTE) and money. Outputs will be the number of participants that are involved in the project.

3.3.2. Data availability and requirements

Data for evaluation can come both from the project and from external sources, but in terms of availability, data from the project is clearly easier accessible than some external sources. Three main steps are necessary to establish the data requirements. First is to establish the units of analysis and the boundaries that are related to it. Second is to link the data requirements to evaluation methods and finally a matrix of analysis can be prepared [2]. This section will look at the first step, while the section on methodology will look at matching the methodology to the data requirements. The units of analysis for evaluation are twofold, based on the difference between monitoring and impact/outcomes measurement. Where the majority of monitoring focuses on the inputs and outputs of projects, the unit of analysis will be the service provider that delivers the intervention, to monitor the resources that the service provider puts into the project. This is data that should be made available by the service provider, but should be easily available and therefore this should be part of the negotiations that shape the evaluation at the start of the project. Secondly is the data that is required for the outcomes and impact, for which the unit of analysis will always be the participant of the project. Data on the indicators related to outcomes will inherently available for all participants, in order to determine outcome payments, therefore data availability is not an issue. Data requirements are more strict on outcomes, especially the ones that are not directly quantified such as behaviour in the education example above. It is important that data collected on these items are done without bias and therefore not by one of the directly involved stakeholders, but that is discussed in the data collection chapter.

3.3.3. Methodology

The choice of methodology is dependent on several factors, among which the evaluation criteria, time and resource constraints, data availability and requirements and finally the desired evaluation rigour [2]. The evaluation criteria are established earlier and have led to a construct for indicators and data requirements. Therefore, the other aspects that need to be considered are that of time and resource constraints. Time and resource constraints are mainly due to the availability of financial resources for a project. The larger the financial resources, the more time and resources can be released for evaluation. Although projects can vary in size, total costs should be assumed for the development of a generic tool. First the potential financial constraints for different project sizes will be determined. Based on a limit of 4% of the total project costs, the evaluation constraints would be €4K for a project costing €100K, €40K for a project of €1m and €400K for a project of €10m. When translated to FTE, based on a normal salary in the Netherlands of €33K, this would translate to 0.12FTE (228 total hours based on 1880 hours for a full FTE), 1.2 FTE (2,280 hours) or 12FTE (22,800 hours). A project has at least two moments of evaluation, pre and post the intervention. Monitoring should take place throughout the intervention and require regular data collection and analysis. If an assumption is made for monthly monitoring moments, a project that has a duration of two years and is split in two cohorts of twelve months, will see 24 monthly monitoring moments and 2 pre, and 2 post meetings, for a total of 28 measuring points. Therefore, based on a project of €100K, this would only allow for 8 hours per evaluation/monitoring moment. When all aspects of collection, processing and analysis are considered this amount is very low. A project of €1m allows for 81 hours per evaluation/monitoring moment, which appears to be enough to work without very strict constraints and therefore allow for a more extensive evaluation tool. The €1m benchmark will be used for further development.

As established previously is working in an quasi-experimental environment, where a control group is used for each project, very costly and time consuming. Working with the methodology of randomised control trials, might provide an optimal environment to measure causality [90], but it might be too costly and project contain so many exogenous factors [54] that this methodology might be too extensive for the goals that it tries to achieve. Also, practically it will limit the potential scale and participation of a project. Therefore collecting baseline data and comparing to that will be the most rigorous method that appears to be within time constraints. This results in a combination between a matched comparison methodology and a reflexive comparison as set out by Adam (2006). Inputs, outputs and outcomes will be evaluated quantitatively, while measurement on impact should be both quantitatively as quantitatively to evaluate processes as well as well-being directly from participants.

3.3.4. Data collection

Data collection will consist of different methods and will be in line with the data requirements. This section will be split in the collection of baseline data and then the collection of further evaluation data, which will be executed separately.

Baseline data can come from different sources. First of all, the baseline data for the well-being of participants of the project, can be measured by a direct questionnaire to those participants, for example through a questionnaire as used by Anand (2006). Baseline data on different indicators can be collected either through collecting baseline data from established institutions. Using the education example from earlier, national or local statistics (whichever is most relevant), can be collected for employment and educational attainment for comparable groups. Looking at educational attendance and behaviour it will be harder however to collect baseline data through available statistics and it is necessary to collect baseline data directly from the schools that the participants are educated at. Therefore collecting baseline data will initially be a mixture of a questionnaire to each individual participant and collecting empirically available data for comparable groups at the start of the project. Finally, when possible, baseline data should also be collected for the inputs and outputs that should be expected in the project. Ideally this will be done directly from the involved service provider, based on their historical inputs and outputs, but this can also be done based on an existing track record. If the service provider does not have an existing track record, empirical data should be collected from comparable service providers. The collection of this last piece of baseline data is something that will be crucial for any project trying to achieve social impact, as costs of the project will be estimated beforehand and will be based on the expected inputs that are necessary to achieve outputs/outcomes. Usually it can be expected that the better and more relevant the track record of a service provider, the more accurate the assumptions around project costs can be.

The collection of monitoring data throughout the duration of the project will be done directly through the service provider, who should report on their inputs and the direct outputs on a regular (assumed monthly) basis. This data can be assumed to be factual as organisations should be required to have their books assessed at least annually and therefore will be assessed. The collection of outcomes data will be done quantitatively through relevant stakeholders. This will be highly dependent on the type of project, but in the educational example, direct collaboration with a school is necessary to collect data on the relevant indicators. Therefore agreements should be sought with relevant stakeholders that are crucial in the data collection process, before the start of the project. For indicators that can be collected from factual results, such as educational attainment, very little subjectivity can be expected. Some of these indicators might actually be subject to judgement, which risks biases in the evaluation. Even though it will be hard to avoid biases from the teachers who fill in the results on behaviour, service providers might have incentives that can skew the outcomes and therefore should not be involved in the data collection of those indicators. Because collecting data by different organisations can become highly complex, all data on outcomes should be collected by a third party evaluator to reduce the risk of bias in the data collection process. Finally, the collection of data on the well-being of the participants will be executed the same as for the baseline data to compare those results appropriately. The process of collection should be identical to allow for comparison, apart from an extension of open-ended questions to assess the process of the project. Data collected on well-being should also be collected by a third party and the questionnaires should be self-administered to reduce the risk of bias [84]. Since the questionnaire of this thesis might touch upon subjects that can be identified as being sensitive, such as income or violence, the method of administering the questionnaire as well as the questions themselves should be fit for that purpose. Not taking sensitivity into account could lead to answers that are less reliable due to social desirability [84]. Sensitive questions can mainly affect the response accuracy. Several studies report higher reporting of drug use for self-administered tests versus methods using an interviewer [84]. There is little effect of computerized versus non-computerized tests, but the effect of having to report answers to an interviewer seems to make the biggest difference, although the interviewer being present in itself has not much effect on the results [84]. The degree of privacy matters, therefore the questionnaire should be administered without the interviewer knowing the answers or who has filled the answers in. To support that, interviews can be taken anonymously and should be handed in a close envelope (or submitted directly online), when participating in the questionnaire.

Collection of data through surveys should be carefully tested before rolling out on a larger scale. Testing of the survey should be done to a small sample of the relevant population, in this case of the service users of the SIB, and the surveys should be conducted under conditions close to those of the main survey [40]. In cases where testing would identify issues with the survey, it should be adapted and with significant changes another test might have to be done to identify whether the changes have helped or potentially created new problems. For multiple reasons, such as efficiency but also of incorporating the expertise of previous studies, a questionnaire design is helped by building upon previous related questionnaires [40].

3.3.5. Data analysis

Collected data should be analysed for causal effects and to determine the financial effects that should follow the achievement (or not) of outcomes. As this process will contain multiple independent variables, the indicators, it is best practice to use a multivariate regression to analyse the variables in the evaluation tool [2]. A regression analysis will present the factors that are most likely to have caused an outcome, therefore providing input into the success of the project and the effects that have caused that. In line with agreements made for that specific project, the data on outcomes and impact should be compared with baseline data in order to determine results based payments for the project. This should be done by straightforward comparison of the difference between actual outcomes and baseline outcome expectations and based upon the financial agreements in line with those results. For further data analysis, potential external effects from the causal diagram, as well as direct inputs should be included to perform a more extensive regression analysis. The outcomes of this analysis can serve as an assessment to determine additional learnings of the project for future implementation, as well as an assessment of the actual outcomes to serve as an improvement of future outcome setting and valuation. Even when it shows significant correlation with external effects, it will be hard to include this in the current project for the financial payments, as this will change cashflows and financial results for organisations in hindsight, potentially causing significant problems for those organisations.

3.4. Conclusions of the criteria for an evaluation tool

Based on the developed criteria, this section will provide conclusions and a summary of the requirements that a generic evaluation tool should meet to appropriately evaluate SIBs. The previous sections have developed the considerations that should be taken when developing such an evaluation tool. When taking all those considerations into accounts, the following requirements are developed:

- Both quantitative and qualitative evaluation

A mix of quantitative and qualitative evaluation will be a strong method of evaluating SIBs. Where quantitative evaluation will be indispensable towards outcome payments, it is important to also understand the processes of a SIB and use qualitative evaluation to get deeper insights and learn from SIBs. Especially in the early stages that SIBs currently are in, using learnings to further improve services is even more crucial.

- Separate evaluations that have service users and service providers as units of analysis

To evaluate both the processes and the impacts, two separate units of analysis should be used. Outcomes and impact have effect on the service users and therefore, evaluation should be performed with them as the unit of analysis on an individual level. However, when evaluating the processes or throughputs, the focus is on what the service provider is actually doing, and more importantly, how it is doing that. This should therefore be evaluated with the service provider as the unit of analysis.

- Outcome indicators should focus on objective and agreed upon criteria

When evaluating numbers, it is important for validity that of the evaluation that those numbers are collected without bias. Whom the burden of proof lies, is mentioned further down, but in each case, effort should be done to minimise the risk of bias both in data collection and analysis. To avoid subjectivity, indicators should be constructed in such a manner that they are based upon agreed upon criteria that are clearly measurable such as employment numbers or school attendance. When this is done properly, the collection will be more objective and valid.

Objectivity is tricky though. When is an outcome really objective? Based on the definition of objective by the Oxford Dictionaries, objective can be described as *(Of a person or their judgement) not influenced by personal feelings or opinions in considering and representing facts*. This means that the indicators should be constructed in a way that they directly represent facts and are not subject to opinions. Even when employment can be seen as partially subjective (i.e. how did you get to employment, and what counts as employment), indicators should be constructed in a way that even in those scenarios they remain without influence by personal feelings or opinions and are agreed upon by all stakeholders.

- Monetisation should not be part of the evaluation

The valuation of outcomes should either be part of the negotiation between the stakeholders [28] or done by the commissioner [90]. But even when it is part of the negotiation between stakeholders, in the end it should be set by the commissioner. The valuation can be linked to the evaluation to come to outcome payments, but is not directly part of the evaluation.

- Effects on society and on people's well-being should be evaluated

When the unit of analysis is the individual who participated in the SIBs intervention(s), evaluation should look at the bigger effects on society (i.e. by looking at the change in employment rates), but also on someone's individual well-being. The latter can be done by the Capabilities Approach, and will include the participant's perception on the effects of the intervention. Evaluating the well-being effects can also point towards the unintended outcomes that might come out of a SIB as displayed in figure 1.3. Unintended outcomes, both positive and negative, would have a direct effect on the well-being of service users. Therefore the well-being should also be evaluated very specifically so that effects on different aspects of people's lives can be seen.

- Evaluation uses baseline data and ex-ante, ex-post evaluation

Using control groups and/or randomised control trials will deliver the highest validity possible. It also brings problems though, in an environment where complete causality is not likely to be proven, one should question whether that should be the top priority. Especially when using control groups also might cause ethical considerations by excluding people from the intervention. In such an environment, the ethical considerations should outweigh the causality in terms of priority. Nonetheless, evaluation should try to prove causality as much as possible. When outcome indicators are based on objective and available data, as stated above, it should be perfectly possible to use baseline data for those indicators. This can be combined with an ex-ante and ex-post evaluation for this specific group and for the baseline data. By evaluating baseline data before and after the intervention as well, it will be tried to include factors such as exogenous shocks on the outcomes.

- Monitoring and subsequent discussion of the monitoring results should take place at least every three months and in between each cohort of a project (when applicable)

Evaluation should be part of a negotiation process, especially when looking at the monitoring of a project. This should be performed frequently, the shorter the feedback loops are, the quicker adaptations can be made when that proves necessary. Shorter feedback loops and more frequent monitoring can also increase costs and raise inconvenience if all stakeholders have to be involved and present for the negotiation. The actual frequency can be different for each SIB, but to have short enough feedback loops, every three months appears to be a minimum for monitoring. When interventions use subsequent cohorts of participants, that will also be a good moment to reflect on the learnings of the previous cohort and overthink the project for future cohorts.

- Evaluation should be performed by a third party

SIBs struggle with the different incentives for all stakeholders, these incentives might drive a bias when the burden of proof lies upon one or more of the stakeholders. Risk of bias and even the perception from the outside world of influencing outcomes should be avoided at all costs. Therefore, evaluation should be performed by third party organisations. This might increase costs because potential synergy effects are lost, but costs should not rise extremely through external organisations.

It is clear that a third party will often need to be paid for its evaluation services, therefore it can be argued that it will never be completely independent. Even though there are little incentives towards negative outcomes, the different stakeholders that are involved, all have different perspectives and therefore different types of incentives. Therefore it will be important that the evaluator is paid by the SIB as a whole, rather than by an individual stakeholder. The combination of the independent evaluator and the objective outcome indicators should increase the validity of the evaluation and make sure that a true representation of the SIB is given, rather than a biased one.

Capabilities

Based on the criteria of an evaluation tool for SIBs, it is important to evaluate the well-being of the people that participate in SIBs. Also, the well-being has an opportunity to at least show the existence of unintended outcomes when measured appropriately. To do so, an evaluation method should be used that can first of all appropriately evaluate the well-being of individuals. Ideally however, the well-being should also be split into different aspects of well-being in order to see on which parts of someone's life the SIB has had effect, and whether that matches the expectation. Hence, looking whether the changes in well-being fall into the intended or unintended outcomes category. An example of that is the Capabilities Approach. In the Capabilities Approach, various aspects of well-being are considered, without weighing them or valuing them.

This chapter will dive deeper into the theories and literature concerning the Capabilities Approach first developed by Sen, but also much developed by amongst others Nussbaum. The section will discuss the available literature about the Capabilities Approach, the pros and cons of the approach and the operationalisation and evaluation examples that are found in existing literature.

In the book *Commodities and Capabilities*, Sen introduces his Capabilities Approach for the first time. More on that approach will be written in the next chapter, but interesting for defining capability is that Sen not necessarily determines capability as something that is a direct physical or mental ability but rather calls it functional capabilities. Emphasizing what someone can do or can be instead of deriving someone's capabilities based on utility, what Sen perceives as wrong.

In his theory on linking capital to capacities, Houghton Budd uses the term capacities rather than capabilities [37]. In that book he uses the concept of associative economics and its view of how capital is based on social aspects. The capacities in his book are predominantly focused on the financial literacy of youth, but do appear to point at the level of financial literacy (or capacity) the youth has. The general definition of capabilities by Sen is that "the capability of a person reflects the alternative combinations of functionings the person can achieve, and from which he or she can choose one collection [68, p.271]. Nussbaum defines capabilities as the answer to the question What is this person able to do and be? [49]. This requires defining some other terms before continuing with the Capabilities Approach. The functionings are beings and doings, where beings are the states that human beings can be in while the doings are the activities that a person can undertake. Another term that plays a key role in the Capabilities Approach is that of real freedoms, or just freedoms, which can be defined as the possibility to achieve alternative functioning combinations [66]. To put these three in a less formal example it could be phrased as the freedom of doing or enjoying something worth doing or enjoying.

4.1. The Capabilities Approach

Since Amartya Sen first proposed the Capabilities Approach in his book *Commodities and Capabilities* in 1985, the approach has undergone a long journey. In this journey it gained both followers and critics, but the approach itself also developed through time. Many articles and books have been written based on the approach, not the least by Sen himself, but another well-known author on the topic has become Martha Nussbaum who's most recent and known contribution is the book *Creating Capabilities : The Human Development Approach* [49]. That last book also became the start of the Human Development

Index of the United Nations . This chapter will try to grasp part of that journey and give an overview of capabilities in the lights of Sen and Nussbaum.

When looking at the bigger picture, the Capabilities Approach can generally be seen as an approach to the well-being of people. A person's well-being can be defined by two types of well-being, well-being achievement (the achievement of functionings) and well-being freedom (a person's freedom to live and be well), which are interrelated, freedom may have importance for well-being achievement [68]. It argues that capabilities should not just be used as an utility function, but should be based on substantive freedoms, on people always having the freedom of choice for people to achieve outcomes that they value. Therefore capabilities in the light of Amartya Sen can be seen as a person's real freedoms or opportunities to achieve functionings [60].

Sen used capabilities for the assessment of human beings to measure their well-being. Important is the fact that well-being isn't merely measured by a utility function or maximising one's well-being. It is possible that a choice is made to do something that does not maximise his or her own well-being, but will improve the well-being of others. The ability to make that choice still contributes to the person's capabilities. This differs from economic theory where choices are made based on the values they represent and a choice should be made to choose the option that has the highest value, to maximise the economic utility of a choice. But if a person has to make that specific choice that would no longer be in line with the concept of freedom of Sen. If we look at more specific economic theory, neoclassical or Austrian economics prescribe that agents will behave solely in its own self-interest, while key to Sen's Capabilities is that a person does not necessarily have to do that. Since the first introduction by Sen, the theory has been taken up by many more economists and philosophers for further development. The most notable of those is the political philosopher Martha Nussbaum. Through collaboration with Sen, but also by her own publications.

4.1.1. Central capabilities

Where the first introduction of Sen contained five components for the measurement of capabilities, Nussbaum eventually introduces a set of ten capabilities [47]. While Nussbaum claims that this set of ten capabilities started independent of Sen by studying the ideas of Aristotle and Marx, they are highly regarded as the next steps of the Capabilities Approach. These capabilities remain arbitrary and can in no way be seen as a fixed set of measurements, as also has been stated by Nussbaum herself, it does give a certain basic understanding of what the Capabilities Approach currently consists of. Nussbaum presents these ten capabilities in her book *Woman and Human Development*:

1. Life; Being able to live to the end of a human life of normal length; not dying prematurely, or before one's life is so reduced as to be not worth living.
2. Bodily Health; Being able to have good health, including reproductive health; to be adequately nourished; to have adequate shelter.
3. Bodily Integrity; Being able to have move freely from place to place; to be secure against violent assault, including sexual assault and domestic violence; having opportunities for sexual satisfaction and for choice in matters of reproduction.
4. Senses, Imagination, Thought; Being able to use the senses, to imagine, think, and reason—and to do these things in a "truly human" way, a way informed and cultivated by an adequate education, including, but by no means limited to, literacy and basic mathematical and scientific training. Being able to use imagination and thought in connection with experiencing and producing works and events of one's own choice, religious, literary, musical, and so forth. Being able to use one's mind in ways protected by guarantees of freedom of expression with respect to both political and artistic speech, and freedom of religious exercise. Being able to have pleasurable experiences and to avoid non-beneficial pain
5. Emotions; Being able to have attachments to things and people outside ourselves; to love those who love and care for us, to grieve at their absence; in general, to love, to grieve, to experience longing, gratitude, and justified anger. Not having one's emotional development blighted by fear and anxiety.
6. Practical reason; Being able to form a conception of the good and to engage in critical reflection about the planning of one's life.

7. Affiliation

- (a) Being able to live with and toward others, to recognize and show concern for other humans, to engage in various forms of social interaction; to be able to imagine the situation of another.
- (b) Having the social bases of self-respect and non-humiliation; being able to be treated as a dignified being whose worth is equal to that of others. This entails provisions of non-discrimination on the basis of race, sex, sexual orientation, ethnicity, caste, religion, national origin and species.

8. Other Species; Being able to live with concern for and in relation to animals, plants, and the world of nature.

9. Play; Being able to laugh, to play, to enjoy recreational activities.

10. Control over one's environment

- (a) Political. Being able to participate effectively in political choices that govern one's life; having the right of political participation, protections of free speech and association.
- (b) Material. Being able to hold property (both land and movable goods), and having property rights on an equal basis with others; having the right to seek employment on an equal basis with others; having the freedom from unwarranted search and seizure. In work, being able to work as a human, exercising practical reason and entering into meaningful relationships of mutual recognition with other workers.

It should be noted that generally when measuring along a scale of multiple ratings - or in this case capabilities - strong ratings can usually make up for weak ratings. In football, a goalkeeper may lack in his tackling ability but makes up for that in his jumping strength which is much more needed as a goalkeeper. Nussbaum argues that in her view, this cannot be the case for Capabilities. The capabilities have no relative weight and cannot be traded off against each other. Therefore, to score well in terms of well-being, one must score good in all ten of the central capabilities. Although the capabilities do not exclude or make up for one another, it is interesting that Nussbaum herself in later writings underlines two capabilities to play a *architectonic* role, affiliation and practical reason [49, p. 39]. She argues that those two capabilities are basically spread out through all other capabilities. People can be in good health, but if they cannot have full practical reason they are still not able to make their own choices freely for other capabilities. The same will go for affiliation. One needs to have the opportunity to engage with others as one wants to while being respected as a being but also with dignity. If a person does not have that possibility, it won't be able to satisfy the other capabilities either.

Now we have defined capabilities and how they are assessed, it is important how to acquire or achieve capabilities. It starts with the notion that the Capabilities Approach does not make a difference between people's innate intelligence or skill, but claims that the goal for all human beings in a nation should be that everyone should get above a certain threshold level of combined capability. In that case, the people who need more help to do so, should also receive more help [49]. Freedom should be taken into account too, just improving someone's functionings (or skills) is not enough, capabilities will only be acquired when this coincides with the freedom to choose whether to use those functionings. In practice, forcing a healthy lifestyle or employment upon people would not increase their capabilities. The aim should be to increase people's freedom and ability to choose the lifestyle they want to live and as long it was a fully free choice, being unemployed or not eating fruit does not need to be contrary with Capabilities. This is an important notion towards the concept of SIBs, forcing people to participate in a programme or even into a job would not improve their capabilities, helping them to gain the ability to make decisions about employment or the location they want to receive end of life care does.

To properly compare the aims and Values of SIBs with the Capabilities Approach, all individual capabilities should be analysed in more detail. Each SIB might have a different goal and if their goals actually align with the improvement of capabilities, it could be that those goals also align with different capabilities.

To analyse capabilities in empirical research, operationalisation of the concepts is needed. This operationalisation has brought up issues as in the definition of the capabilities approach there is no clear statement on how to apply the capabilities in empirical analysis [35]. Previous findings range

from statements that argue that it is radically underspecified how the Capabilities Approach might be applied to various applications in empirical research [6, 10, 12, 63]

Important in operationalising capabilities is the earlier mentioned distinction between functionings and capabilities. While the functionings can be measurable as they are what people do and are therefore clear outcomes (i.e. work), the capabilities assess the freedom to realise specific functionings [64] and that freedom is harder to measure.

It is important that the different parts of the Capabilities Approach are distinguished and public policy focuses on the combination of how the resources are transferred into capabilities, thus the conversion factors [94]. The outcome of that conversion are capabilities and as of then it is up to the freedom of choice of a person, whether a person is able to make a completely free choice towards the conversion into functionings.

Different studies have suggested a framework for researching the capabilities approach and all have in common that the right capabilities should be selected that are relevant to a specific subject. Examples are that of a mixed methods approach where core dimensions and participatory methods are used [4], a two-step framework of deriving a core capability list from the international human rights framework and then supplementing and refining that core list through democratic deliberation and debate [95, 96]. Finally a list of five criteria that the selection of appropriate capabilities should always satisfy [59, p.70-71]:

1. The criterion of explicit formulation: the list of capabilities needs to be explicit, discussed and defended
2. The criterion of methodological justification: the method used to generate the list of capabilities needs to be justified
3. The criterion of sensitivity to context: the list of capabilities should be context specific/sensitive
4. The criterion of different levels of generality: lists that will be applied empirically should be generated in two stages, (i) drawing up an ideal list and (ii) then drawing up a pragmatic list
5. The criterion of exhaustion and non reduction: lists of capabilities should include all important elements

Even Nussbaum and Sen are not in agreement between the use of a fixed list of capabilities as for instance the list of ten central capabilities [47]. This can be seen by the arguments in favour of a set list, saying that on list of central capabilities is necessary for the Capabilities Approach to offer a valuable normative gender perspective [48] and the arguments against such a method that state that although core capabilities might be defined, an predetermined and exhaustive list of capabilities would not meet the variety and complexity of the subjects that might be analysed using the Capabilities Approach [67].

It can be argued however that whether it is a fixed set of capabilities or a list that is specific to the context of the subject, to assess capabilities empirically there an a priori definition of a set of basic capabilities needs to be established [21].

Another example of a list of ten central capabilities [96, p.5-8] shows that although coming up with a fixed list appears to be a useful exercise and give overview, the subtle differences between this list and the list that was defined by Nussbaum also show that these central capabilities not necessarily cover every social issue:

1. The capability to be alive
2. The capability to live in physical security
3. The capability to be healthy
4. The capability to be knowledgeable, to understand and reason, and to have the skills to participate in society
5. The capability to enjoy a comfortable standard of living, with independence and security
6. The capability to engage in productive and valued activities
7. The capability to enjoy individual, family and social life

8. The capability to participate in decision-making, have a voice and influence
9. The capability of being and expressing yourself, and having self-respect
10. The capability of knowing you will be protected and treated fairly by the law

This list of capabilities definitely has different aspects than that of Nussbaum, but was recommended to use for the measurement of inequality in Britain [96]. However, both lists leave room for interpretation, it is so broad that services to help the people who need it [49] can hardly target, let alone improve, all of those capabilities at once. This is why Sen also argues that capabilities should be selected in context of the study itself and the Capabilities Approach should be more used as a framework of analysis [59]. The lack of need for a list of core capabilities is shown in the goals that the Capabilities Approach is used for, as one cannot understand the values of all the people in the world and the capabilities selected need to be unbiased and legitimate [35]. This is why Alkire [4] argues that analysis for cases should be context specific and therefore there should not be a single list of capabilities.

In performing empirical research based on the Capabilities Approach there are five issues to consider [42, p.5]:

- The adequate evaluative space: capability vs. (achieved) functioning
- A list of (essential, relevant) capabilities or functionings
- A set of indicators relate to the selected dimensions of well-being and adequate criteria to measure and represent them
- How (and if) to aggregate the elementary indicators to obtain an overall evaluation for each single dimension (functioning/capability) of well-being
- How (and if) to add up all the dimensions and to reach an overall evaluation of well-being

Based on this, Martinetti decided that basing analysis on functionings is a more practicable method than basing on capabilities.

Several studies have performed analysis using the Capabilities Approach through either a quantitative or qualitative method. These either use the approach of Sen and merely use the Capabilities as a framework or use a predefined list of central capabilities and try to operationalise/measure accordingly. Anand et al [6] made use of the ten central capabilities constructed by Nussbaum to measure well-being in British households. In that, questions around the capabilities approach are made that distinguish functionings and capabilities, resulting in an analysis of the impact of each of the ten central capabilities on people's satisfaction.

Anand was part of another empirical study based on the Capabilities Approach [5]. Here it identifies two distinctions in the assessment of well-being made by Sen himself [65], which distinguishes one's goals from well-being and achievement from freedom to achieve. Therefore, Anand and Van Wees identified four categories to measure. The study also recognises limitations of focusing just on capabilities, as people may not realise that certain functioning may be achievable. Also there can be a difference between the functionings an individual thinks it can achieve and what it actually can achieve [5].

Different strategies are discussed to apply the Capabilities Approach are discussed by Brandolini and Alessio [12]. The focus of this research lies more on functionings and assesses the health, education, employment, housing, social relationship and economic resources through a survey.

4.2. Critique on the Capabilities Approach

Apart from various research and supporting papers for the Capabilities Approach, there is obviously also critique. Especially Amartya Sen is highly regarded as an economist, shown for example by the Nobel Memorial Prize in Economic Sciences that he received in 1998. While Nussbaum is a philosopher rather than an economist, the Capabilities Approach should still be treated as an economic theory and most critique actually looks at its economic values. Sen himself argues that his studies are mostly formed by the work of Adam Smith and even goes back as far as Aristotle. Especially the latter is very much embedded in practical reasoning, which can be argued to be in line with current mainstream economics.

As can be seen by the various additions and maybe even variations on the Capabilities Approach, the start of critique can be found in the fact that the Capabilities Approach as initially proposed by Sen was incomplete. Something that even Sen acknowledges [65]. It is probably argued that the Capabilities Approach is more a micro economic theory than a macro theory since it is very much focused on the status of an individual. There are obviously some similarities with regular micro economics and it could be argued that the Capabilities Approach also very much looks for maximum utilisation. The definition of utilisation however is different as the capabilities are not monetised where that monetisation lies at the basis of mainstream economics. As discussed, SIBs also mainly looks at monetisable outcomes and this is where the Capabilities Approach could potentially add value.

In contrast to mainstream economics, the capabilities approach implies that neither opulence nor utility adequately represent human well-being and deprivation [20]. There is for instance a clash between Nussbaum's stand and that of utilitarianism, where utilitarianism allows for the idea to give weight to adaptive preferences, Nussbaum's view on the Capabilities Approach rejects this and states that no capability can be seen as more important than another [78]. There is obviously a discrepancy between the theory of the Capability Approach and for instance that of Neoclassical economics which bases on the free market principle, full rationality and optimising (financial) utility. However if you look at different descriptions of Neoclassical economics, it could be argued that more modern day neoclassical economics actually assumes that individuals are not always optimising financially, but are optimising their preferences which can take into account the preferences of others and/or is open to adapting preferences [9]. This view of neoclassical economics would definitely lie closer to the Capabilities Approach. Finally, it could be argued that the Capabilities Approach lies closest to the ideas of Keynes. A concept as striving for full employment can be considered something that is in line with the Capabilities Approach, but also Keynesian economics focuses very much on items such as consumption, investment and aggregate demand.

The main critique on the approach however lies in the way it is operationalised. This partly originates in the idea that Sen has no concrete list of capabilities, but this has been raised but also expanded on by Nussbaum and Vizard [47, 96]. Even with these lists of capabilities, issues are raised around the way capabilities are being operationalised and measured. Measuring well-being is something that can be hard and is often perceived as something that can't be done, however it is something for which there is also supporting evidence [50]. As the OECD recognises the measurement is highly sensitive to measurement methodology, which also makes for tough comparisons of these types of measurements with other research or from even other researches. Therefore reservations should remain around the measurability of the Capabilities Approach, but for each research, sound methodology should be developed and consistently applied to create reliability in data.

4.3. Operationalising individual capabilities

As discussed in the previous sections, there is some difference of opinion about the use of a specific list of capabilities or just using capabilities as a framework. Also there is critique on the ability to make the Capabilities Operational [20]. This research will use the capabilities created by Nussbaum [47] and based on surveys and empirical research that has already been conducted [5–7]. This will serve as a basis for the later selection of criteria to question in a survey in which indicators will be selected based on the six steps that are suggested by Alkire [3, 61]. Furthermore it also looks at examples of how the Capabilities Approach is operationalised in previous research.

1. The functioning belongs to the capability set (is itself valuable) OR the functioning is directly associated with the capability set (highly correlated, etc)
2. The functioning pertains to a basic human need, i.e. that without which one's life may be blighted
3. The functioning is not significantly dependent on any non-basic prior functioning
4. The functioning is not dependent on the presence of uncommon ability or interest
5. A level of achieved functioning which is widely recognised to be basic can be specified and empirically observed
6. Provision of the functioning does not necessarily compromise freedom to pursue other significant functionings in the long term

To start, all ten central capabilities are analysed for methods of operationalisation and questioning.

1. Life

The capability of Life is described by Nussbaum itself as *Being able to live to the end of a human life of normal length; not dying prematurely, or before one's life is so reduced as to be not worth living* [47]. It can be assessed that there is a U-shaped pattern on someone's age and its overall satisfaction in life [6, 11], where there is a low point around the age of 40. Other data suggests that life satisfaction increases with age [23]. This all points to the capability as described by Nussbaum that being able to live to the end of a human life of normal length indeed is a useful capability.

2. Bodily Health

Nussbaum describes the capability of Bodily Health as *Being able to have good health, including reproductive health; to be adequately nourished; to have adequate shelter* [47]. Especially with health there is data that poor health restricts well-being. Health that limits someone's ability for daily activities reduces overall satisfaction [6], which is also in line with the definition of capabilities itself where it is what someone can do. Veenhoven [92] also finds that people who are in good physical shape tend to be happier. Brandolini and D'Alessio proposed indicators such as the ability to walk 100 meters and contacts with doctors and nurses [12].

Interestingly, the operationalisation of the capability of Bodily Health usually is done by using the aspects of Life, i.e. life expectancy [61].

Additionally Anand et al also find that being well nourished (there defined as eating meat, chicken or fish, every second day and if not, whether this is because they can not afford to) has a positive effect on overall satisfaction. They have also measured having adequate shelter as being able to move if they want to move house, but although there was a positive effect on satisfaction with partner and job, the effect on overall satisfaction was found negative. Both are confirmed by [58, 61].

3. Bodily Integrity

Bodily Integrity is described by Nussbaum as *Being able to have move freely from place to place; to be secure against violent assault, including sexual assault and domestic violence; having opportunities for sexual satisfaction and for choice in matters of reproduction* [47]. The ability to freedom is defined by Anand et al as the option to make use of a car or van, but only had a positive effect on men. Also it made use of the British House Panel Survey's data on vandalism or crime in the area, but couldn't identify whether that includes sexual assault and domestic violence [6]. Other data however suggests a strong correlation between happiness and murder rates/lethal accidents in a country and that the acceptance of homosexuality and prostitution was also strongly correlated with happiness [92]. Exposure to violence and thefts is described as a proxy as well by Brandolini [12]. Other variables that are mentioned are a non-hazardous work-environment, non-hazardous physical environment and safe birth control and child bearing [61].

4. Senses, Imagination and Thought

The fourth on the list of central capabilities has a somewhat elaborate explanation by Nussbaum: *Being able to use the senses, to imagine, think, and reason—and to do these things in a "truly human" way, a way informed and cultivated by an adequate education, including, but by no means limited to, literacy and basic mathematical and scientific training. Being able to use imagination and thought in connection with experiencing and producing works and events of one's own choice, religious, literary, musical, and so forth. Being able to use one's mind in ways protected by guarantees of freedom of expression with respect to both political and artistic speech, and freedom of religious exercise. Being able to have pleasurable experiences and to avoid non-beneficial pain* [47]. The core of this capability appears to be education, which is also the most easily measurable variable based on the attainment of education [12, 58, 61], which is in line with the capabilities distinguished by Sen [64]. Although education appears to be related to higher job and income satisfaction, overall data suggests that the attainment of higher education actually causes lower overall satisfaction [6]. This is argued to be caused by less appropriate jobs [92] or by related higher expectations. However, overall do countries that provide most education appear to have the happiest people [92].

5. Emotions

The capability Emotions is defined as *Being able to have attachments to things and people outside ourselves; to love those who love and care for us, to grieve at their absence; in general, to love, to grieve, to experience longing, gratitude, and justified anger. Not having one's emotional development blighted by fear and anxiety* [47]. This capability can be rather hard to measure and [6] uses questions on not losing sleep through worry and not feeling constantly under strain as proxies. This is both found to have a positive effect on overall satisfaction. Another method that is used is to look at social interaction or primary relationships [12, 61].

6. Practical Reason

Practical Reason is defined as *Being able to form a conception of the good and to engage in critical reflection about the planning of one's life* [47]. The proxy used by Anand et al [6] delivers mixed results on the ability to face problems and its correlation with satisfaction. Two other studies have identified mental health/capacity as a proxy for practical reason, but this might be even harder to measure [61].

7. Affiliation

The capability of Affiliation is split into two sub-items, (i) *Being able to live with and toward others, to recognize and show concern for other humans, to engage in various forms of social interaction; to be able to imagine the situation of another* and (ii) *Having the social bases of self-respect and non-humiliation; being able to be treated as a dignified being whose worth is equal to that of others. This entails provisions of non-discrimination on the basis of race, sex, sexual orientation, ethnicity, caste, religion, national origin and species* [47]. Basic levels of aspiration and self respect are seen to be linked to Affiliation [61]. An example is that data have found that unemployed have in general lower levels of aspiration than employed people [56]. Employment might therefore be a decent proxy for affiliation. There is also positive effects found from being able to go on holiday, have friends or family round and the same goes for people that don't see themselves as a worthless person [6].

8. Other Species

Eight on the list is the relationship not with one's human environment but with nature. It is defined as *Being able to live with concern for and in relation to animals, plants, and the world of nature* [47]. However, in none of the literature that was examined, was there a proxy to measure this empirically.

9. Play

Play refers to being able to engage in fun and recreational activities, defined as *Being able to laugh, to play, to enjoy recreational activities* [47]. Data suggests that being happy obviously results in more overall satisfaction. At the same time, being able to enjoy day-to-day activities more than usual has no effect, but not being able to do day-to-day activities does have a negative effect on overall satisfaction [6]. An indicator could be the number of vacation trips someone takes or the leisure-time pursuits [12].

10. Control over one's environment

The final capability on Nussbaum's list is control over one's environment and its definition is also split in two parts: (i) *Political. Being able to participate effectively in political choices that govern one's life; having the right of political participation, protections of free speech and association* and (ii) *Material. Being able to hold property (both land and movable goods), and having property rights on an equal basis with others; having the right to seek employment on an equal basis with others; having the freedom from unwarranted search and seizure. In work, being able to work as a human, exercising practical reason and entering into meaningful relationships of mutual recognition with other workers* [47]. Interestingly there was no effect of the ability to vote in an election on satisfaction [6], however it is used as a proxy for measurement, combined with the ability to file political complaints [12]. For women there was significant negative effect found on their health limiting their ability to do certain types of work, but not for men [6]. Other data however suggest that there is strong correlation between political freedom and perceived freedom in society. Also social inequality can affect happiness negatively as it may limit people [92]. This can also be seen as having control over all your options, being economically secure to control your life and being physically secure to control your actions day by day [61]. Being economically secure can also be seen as having the ability to cover unforeseen expenses of up to \$1,000 within a week [12]. Brandolini also agrees with holding property as a material capability.

4.4. Measuring well-being

As this research does not intend to compare individuals or even groups of individuals with others, there appear to be less issues around setting up a framework to determine how you can rank people's complete set of capabilities. However, when measuring the full capability set of an individual on two moments to distinguish the difference, it could very well happen that certain capabilities have improved while others have deteriorated. To then establish if the person's has improved or worsened its full capabilities, a method for measuring that difference should be decided upon. There are seven available methods found to do so.

1. *Dominance partial ordering* The first option is to use a method that is similar to that of Pareto efficiency, being that individuals are considered better than another, if the values for all functionings/capabilities are equal or higher than the other individual [61]. This method is confirmed by Sen when he says that, as there might be disagreement about allocating weights to capabilities, an improvement can be seen if there is an improvement in all capabilities. Sen calls this dominance partial ordering [64, 66].

2. *Sequential dominance* This method is so far mainly used to look at income distribution versus a family's needs. It transfers the income itself to how much income the family thinks it needs and ranks these equivalents to each other [10]. Although it is mentioned that this can be used for partial ordering as well [12], it does not appear that more empirical applications than income and needs are used so far.

3. *Multivariate statistical techniques* This technique is used to analyse responses to qualitative questions, although it was stressed that it was more of a data reduction than that it provides full analysis of relative valuations [63]. However it is used to explain patterns of responses from qualitative questionnaires.

4. *Multidimensional inequality or poverty indexes* The advantage of using inequality or poverty indexes is that a valuation function is implicitly defined in the index [12]. Therefore this can be seen as a full aggregative strategy if the index is set up on an individual basis, for example with [85], but also the Human Development Index is something that could be used on a country level [87].

5. *Full aggregative strategy* In the full aggregative strategy, aggregation is taking place without the use of a predefined index. This means that some indicators should be chosen as summary indicators of well-being [12]. This could be indicators that summarize the full set of capabilities, but potentially also could be that of a summarizing indicator for a relevant set of capabilities or functionings.

6. *Borda rule ranking* The Borda rule makes use of ordinal ranking to compare individuals on individual functionings or capabilities and is therefore useful for comparison of people amongst each other. It takes the well-being results of a specific functioning of N individuals and ranks the lowest person a 1 and the highest person N. The individuals can then be ranked for different functionings/capabilities [61].

7. *Composite index* The final method that is identified is similar to the Borda rule ranking, but uses a scalar method. The highest score is determined a 1 and all other scores will be a fraction of that and will therefore have a score between 0 and 1. The average of the scores of a person is then used to compare with other individuals [61].

4.5. Conclusions of the Capabilities Approach

In this chapter, the Capabilities Approach and other approaches to evaluate well-being are discussed to look at its use for an evaluation tool for Social Impact Bonds. It is clear that various methodologies have been used, but all appear to overlap to a certain extent. The Capabilities Approach appears to have gained most acceptance, also shown by the use of it as a basis for the Human Development Index [87]. The approach also has weaknesses though. Using several studies, a way of operationalising the approach has been found, but this remains sensitive as measuring the well-being of people is something that remains complex and will potentially never be done completely accurate. The ten central capabilities that have been proposed cover most aspects and will therefore be a good start towards evaluating well-being of participants. Nonetheless, it can be hard to identify all aspects of someone's well-being as you can't predetermine all aspects of someone's life.

The Capabilities Approach doesn't attach weights to the individual capabilities, which is positive as capabilities cannot be traded off against each other, but it can be questioned whether this will also apply in real life. An individual human being, might rank certain aspects of its life or well-being as more important than others. Moreover, this ranking is likely to change at different points in one's life. This

makes it harder to assess the actual progress of someone's well-being as an improvement on an area that is too lesser importance for the individual, might count as less overall progress than when important aspects of someone's life have improved. For the Social Impact Bonds and its system it remain most important though to evaluate to what extent certain capabilities have improved as it was set out to and whether other aspects of the participant's life were also affected.

It is clear that the Capabilities Approach is widely used to theoretically and empirically assess people's well-being. From the three types of data, market purchase data, responses to questionnaires and non-market observations of personal states [64], this research will focus most on responses to questionnaires. Based on the applications that have been identified, the multivariate statistical technique as used by Schokkaert and Van Ootegem [63] appears to be the application that is most useful for this research and they have provided a questionnaire focused on unemployment that is used for further reference. Also the questionnaires as used by Anand et al, and Brandolini and D'Alessio will be used as a starting point for establishing a solid questionnaire for this research [6, 12].

Characteristics of a generic evaluation tool for Social Impact Bonds

Now that criteria have been set for an evaluation tool, this should be translated in an actual design of the tool. This chapter will move further into the design of an initial design of the evaluation tool, before it can be tested on the case study of Social Impact Bonds. While some of the parts of the evaluation will be specific to a project, others can be set and used generically for all different SIBs. Monitoring and evaluation will be treated separately and both will be developed in this chapter. Finally, also separate parts will be developed for evaluating service users and service providers.

5.1. Evaluation

The first part to look at is evaluation. Although monitoring will be more frequent, some of the evaluation tools will also be used for monitoring. This section will look into the indicators, when evaluation takes place and what it then evaluates. Then it will look at evaluating the impact on, and perspectives of, service users. Finally the evaluation of the processes will be discussed. It should make sure that all parts of the system as described in figure 1.3 are being evaluated. This will therefore be split in an evaluation of the processes of the service provider itself, how is the intervention delivered. In that case, the service provider will be the unit of analysis, but this can be best tested by using the participants as the units of observation, as they can judge on the intervention they have received. The second part is the evaluation of the system of the SIB, did everything work out as initially thought. This evaluation should inform the design of future SIBs based on the learnings of the evaluated SIB.

5.1.1. Evaluating outcomes

Outcomes of a project can be seen as the goals or indicators that were distilled from the causal model of a SIB. As discussed in the previous chapter, those outcomes should be based on agreed upon indicators that can be measured objectively. Also data should be collected through direct data rather than questionnaires or surveys, for instance through collecting unemployment data. This data is dependent on the specific Social Impact Bond, so can't be predefined for all projects. However, data collection methods can be defined. For each cohort of a project, evaluating outcomes should follow the following steps:

1. *Determine criteria for baseline data* The baseline data should be collected so that it represents the group of service participants as closely as possible, while remaining a large enough sample to be statistically significant. This should be based upon the established criteria for the participants of the SIB. When for some reason, this can't be exactly replicated for the baseline data collection, the effects of using different types of groups should be assumed through a negotiation process between the involved stakeholders.

2. *Collect data at start of the intervention for baseline and participating service users* While the characteristics of the baseline group and the participants should be the same, in some cases they will be looking at different times of their life. For instance, when the goal is to improve the exam results of young people in secondary education. It can't be collected what those young people currently achieve,

as they will not have done their exams yet. Therefore, data collection should be focused on how young people with similar characteristics currently (or at the last available moment), performed at those exams and that data will be used as the baseline. In this example, the collected data before the intervention will therefore be only one set of data rather than two. When attendance records are observed however, data for both the participants and baseline data (for instance for all young people at that school in the same year) can be collected. When both can be done, this will lead to a better control for exogenous effects and the results will be more valid.

3. *Collect data after the intervention has completed for baseline and participating service users* Data will be collected similarly after the intervention to before, as described in step 2. However, in some cases this will be based on empirically collected data directly from the participants. Again, when possible data will be collected both for a baseline and for the participants. When this is not possible, only data from the participants will be collected.

4. *Analyse data to determine final outcomes* This step can be executed in two ways, based on the amount and types of data that can be collected. When data for an indicator can be collected separately for the baseline and for the participants (such as the attendance example), evaluation should look at the development of the results for the participants versus the development of the results for the baseline, to compensate for potential exogenous effects. The difference between the developments can be seen as the actual outcome of the intervention and should be used towards outcome payments after the evaluation is completed.

For indicators where baseline and participant data can't be collected separately, the data collected of the participants should be directly compared with baseline data as collected before the intervention. This will determine an *actual outcomes* versus *expected outcomes* comparison which will be used to determine the final results of an intervention. This method is more open to exogenous risks, but falls within the ethical and methodological criteria of evaluation social interventions.

Evaluating the outcomes refers back to two main aspects of the SIB system. It obviously directly evaluates the intended outcomes, as they are based on the indicators as set at the start of the intervention. But when the baseline data is used correctly it also looks at the exogenous effects as they will have an effect on the changes of the overall baseline data. When changes in the baseline data can't be measured, the evaluation will be more open to exogenous risks, but the outcomes can still be controlled for outcomes that would have happened without the intervention as well.

5.1.2. Evaluating impact on participants

Evaluating the impact interventions have on service users can be done by directly measuring the well-being of service users. This can be done by doing an ex-ante and ex-post test of the well-being of all service users. While this method will allow for exogenous effects to influence the outcomes of this part of the evaluation, combining this evaluation with the qualitative evaluation can give more insights in the causality of the intervention. The Capability Approach is a methodology that has previously been used to measure well-being of people. The questionnaire provided by [7], covers the full capabilities of people and therefore is appropriate of measuring the capabilities of all participant. This will allow for comparison on the full spectrum of capabilities, but also for evaluation of the improvement of specific capabilities.

By evaluating the well-being of participants and the impact the service has directly on them, the intended outcomes will be evaluated, because it is tested whether the participants also feel that the change that is intended, actually had a positive effect. Evaluating well-being with the Capabilities Approach also tests whether there have been any unintended outcomes after the intervention. By evaluating the different capabilities, the intended outcomes would have realised changes on a set of predefined capabilities. However, by comparing the effects amongst other individuals and looking at the other capabilities, it can be seen if the participant has experienced other effects than the intended ones.

5.1.3. Evaluating the service provider's processes

This part of the evaluation will focus on the evaluation of the processes with the service provider as its unit of analysis. As the unit of observation though, it will use the participants to give input on the processes of the service. This therefore should be done through a qualitative analysis on a sample of the participants of a SIB to gain deeper insights in the processes of the intervention and the effects of them. Therefore, a questionnaire should be developed and used that tests the processes qualitatively. Even

though the exact interventions might be different, the perception of processes can be tested generically. These processes are crucial as they form the core of the system of a SIB.

The evaluation should build further upon the capabilities approach by asking around causality of the project as well as the processes of the intervention. While the questionnaire of the capabilities approach is quite extensive, and the unit of observation will be the same, the qualitative evaluation should be much more concise. Therefore a set of questions is developed to test mainly those two themes. Questions should be open and have the intention of capturing insights.

Therefore the designed questionnaire will include the following questions:

1. Why did you participate in the intervention? (If someone referred you to the intervention, who?)
2. Do you feel that you have benefited from the intervention (Yes/No), please explain why
3. Could you explain which parts of the intervention could be improved and why?
4. What do you think of the guidance during the intervention?
5. What do you think of the duration of the intervention?
6. What did you hope to get out of the intervention?
7. Which of those goals did you realise thanks to the intervention?
8. Which of those goals do you think you would have realised if you hadn't taken part in the intervention?
9. How could you have benefited more from the intervention?

These generic question can be expanded by a maximum of 3 open questions that are intervention specific. The final questionnaire can be combined with the closed and quantitative questionnaire concerning the Capabilities Approach.

5.1.4. Evaluating the SIB's processes

Evaluating the processes of the SIB as a whole by interviewing the stakeholders can add value and provide valuable insights into the processes of the SIB system and the interactions between the stakeholders. When the SIB is viewed as a process, or even a process within a specific system, this also means that it is important to figure out how that process exactly has been working. Therefore, the evaluation should provide insights into that process. The interview should be qualitatively to gain deeper insights and allow for additional deeper questions during the interview in order to create a full picture of the SIB and its processes. All stakeholders that were directly involved in the SIB should be interviewed and those interviews should be analysed to gain more insights and the processing of that analysis should provide learnings for all stakeholders towards future SIBs.

The questions for the interview are developed to gain those broader insights into the general process, but also address specific topics of that process. It aims to learn more about the individual motives and outcomes of the stakeholder, what aspects of the system have influenced them and how, and finally how their view is on the governance of the system. Those items are further developed into a questionnaire.

Therefore a new questionnaire is designed with the following questions:

1. Why did you participate in this Social Impact Bond?
2. As an organisation, have you benefited from the SIB?
3. Do you think the overall goals of the SIB have been realised?
 - If not, why?
4. Do you think your individual goals for this SIB have been realised?
 - If not, why?
5. How could you have benefited more from the SIB?

6. How have the relationships with the other stakeholders developed?
7. How could the ... have benefited more from the SIB?
 - Commissioner
 - Investor
 - Service Provider
 - Participant
8. How would you have made the agreements with the other stakeholders differently than set out in this SIB, using what you have learned in this SIB?
9. What outcomes were the result of this SIB that you hadn't foreseen at the start?
10. In hindsight, what changes would you have made to the indicators that have been used for this SIB?
11. Overall, could you explain which parts of the SIB could be improved and why?

These generic questions should give a broader insight into the process of the SIB and can be used to provide learnings for the future of SIBs. For question 7, the questions should be asked for all of the stakeholders apart from the stakeholder that is being interviewed at that moment. This is to gain more insights in the perceptions of all stakeholders into the others of the SIB.

5.2. Monitoring

While monitoring overlaps somewhat with the evaluation, some evaluation processes are different and especially the timing is different. This section will look into the design of the monitoring aspect of the evaluation tool. Monitoring will take place at least quarterly, but when budget or stakeholders time allows it, this could be more frequently. The aim is to involve all stakeholders in the processes of the intervention, learn from already performed actions and continuously improve the services. When monitoring is done correctly, it can lead to adaptive decision making within the SIB. Insights gained from monitoring will result in a frequent feedback loop that can be used to adapt to changing circumstances. This can be changing external circumstances, but also internal changes. When monitoring provides new insights based on which it is thought that changing certain processes will lead to better results, this should be done, therefore leading to adaptive decision making based on the monitoring aspect of the tool.

5.2.1. Monitoring inputs and outputs

The most clear part of monitoring, looks at the inputs and outputs of the service. Outputs differ from outcomes, but measure the number of participants that take part in the service, or in specific parts of the service when applicable. It should be included what the expected outcomes are based on the outputs of the monitoring, to give all stakeholders an indication towards the potential final results. To give a fair indication, this should be done based on the collected baseline data, success rates of previous cohorts of the same intervention, or assumptions of the success rates based on a negotiation process between all stakeholders. All collected information for monitoring should be discussed in a meeting between at least the evaluators, service provider and investors. The commissioner can join in some circumstances, but is not always needed in those meetings. These meetings should provide in discussion of the inputs and outputs, as well as potentially allow for adjustments to the intervention, when deemed necessary. Evaluating inputs can also be done in order to compare the actual inputs with the inputs as they were set out at the start of the intervention. The following information should be collected and discussed:

- All human capital that has been involved in the execution of the intervention (this includes both human capital that was foreseen and/or is being paid for as additional human capital that has not been paid for or is involved on a voluntary basis)
- All financial capital that has been used to date
- All other resources that have been used for the intervention

- A comparison of the used resources and the budgeted resources for the intervention until that point in time
- All participants that are currently participating in the intervention
- When relevant: the stage current participants are in and what part of the intervention they are involved in
- A comparison of the number of participants and the expected number of participants until that point in time
- Projection of expected outcomes based on the number of participants, the stage they are in and the expected success rates
- A comparison of the total expected outcomes as currently standing and the negotiated expected outcomes as set when the SIB started

5.2.2. Monitoring service provider's processes

Inputs and Outputs give vital information into the progress of the SIB and are both necessary to monitor. Just as with the final evaluation though, it is very important to gain deeper insights into the processes intervention in an ongoing manner. Therefore, the developed qualitative questionnaire should be also used throughout the process of the intervention. It is not feasible to use the questionnaire on all participants, as this will be a lengthy and time consuming process which is not applicable for frequent monitoring. It can however be used on a small sample of participants to inform the stakeholders of deeper insights of the service to make better decisions on the next steps of that intervention. The size of that sample can be very small and for the monitoring it should not meet any criteria for statistical significance, the primary goal is to provide additional insights on top of monitoring inputs and outputs.

5.3. Governance of the tool

Aside from the already discussed characteristics and design of the tool, it should also be discussed how the tool is used. This section will therefore go deeper into that aspect of the tool and will discuss who the evaluator should be, which stakeholders should be involved at what point and how decisions should be made based on the evaluation.

As discussed in chapter two. Evaluation should be performed by an independent organisation that is not a direct stakeholder of the SIB. While the costs of hiring the evaluator should be carried by the whole SIB rather than just one stakeholder, in order to avoid bias towards a single stakeholder, the overall evaluation still risks bias towards positive outcomes as the SIB as a whole benefits from a positive evaluation. Several things should be set up in order to avoid that bias and to make sure that evaluation of a third party is truly independent. The first is obviously to use the objective and hard outcomes that were discussed earlier, but that is not directly linked to the governance of the tool. Other aspects to ensure a valid evaluation should be that the fee of the evaluator is never in any way linked to the performance of the contract. Also, the outcomes and the methodology of the evaluation should be transparent. This transparency is important for the stakeholders themselves, as they can see how the outcomes are achieved, but transparency outwards should also be maintained. By allowing the external environment to engage in the results and methodology of the evaluation, this will put pressure on the evaluator as society can directly spot biases. To also reduce the risks of biases in favour of any stakeholder, the evaluations and monitoring should be discussed with all stakeholders and the finalization of the results should be agreed upon unanimously by all involved stakeholders. It remains important to realise though that evaluators can hardly be fully independent. Dependencies can arise in several ways when different stakeholders are involved that all have incentives towards a bias, as discussed earlier in the actor analysis. Even when all measures are taken towards the evaluator, potentially other forms of bias may arise. Through repeated use of the same party or even when the party is dependent on its evaluation to get other assignments in the market. Those effects can't be excluded in current society. In the evaluation tool, this is where not just the governance plays a role but also the structure of the evaluation and the agreements on the indicators that will be used for measurement. These agreements should be strict and transparent to reduce the risk for bias further.

The main role in the evaluation process is obviously for the evaluator. This organisation will be responsible for collecting, analysing and reporting the results of the evaluation. However, this will also involve the cooperation of several other stakeholders, which should therefore be guaranteed when necessary, in order to make sure evaluation takes place properly. The service provider, together with its participants will be involved as the service provider will have to provide data and should work with the evaluator to make sure that participants engage in the evaluation of their well-being and the processes of the service provider. Investors are not directly involved as they have no role in providing data. The commissioner also plays an important role in the evaluation process though. Most indicators will be established based on existing records, but definitely all baseline data will be based on existing databases. As the evaluator is a third party, this might create some legal barriers as to getting access to personal and/or governmental data. Therefore the commissioner, together with the other stakeholders of the SIB, should ensure beforehand that the third party evaluator can get access to the needed data. During the evaluation process the commissioner should also ensure and work with the evaluator to make sure that it will remain to have access to all required data in order to perform a full and valid evaluation. Finally, the evaluator should present the results of the monitoring in frequent (quarterly is suggested) meetings, where the results will be shared with all stakeholders before the meeting, and discussion of the evaluation can take place at the meeting. The same will go for the evaluation itself, but the timing of this will be dependent on the end of a cohort or the full SIB.

While the evaluation process can inform decision making within the SIB or even for future SIB, it could also happen the other way around. Based on the evaluation processes, or developments in the SIB, it could be suggested to change the way evaluation is performed. This could happen in any of numerous situation, so no comments can be made on its contents beforehand. It is critical though, that whatever changes are considered or even made to the evaluation process during the SIB, they should be such that they don't pose threats to the validity of the actual evaluation of the achieved outcomes. It is understandable however, that when evaluations or other aspects provide new insights and it is thoughts that certain aspects of evaluation or, most likely, monitoring, can be altered to provide additional insights to the stakeholders.

5.4. Conclusions of the initial design

This chapter has provided an initial design of a generic evaluation tool for SIBs. It consists of several items, of which some are predefined, while others will be more specific for the SIB that is being evaluated. The tool is split up in an evaluation and a monitoring part. Evaluation will be service specific for the outcomes/indicators as those will clearly differ based on the aim of the services, but a method of evaluating the specific outcomes is set out. Other parts of the evaluation tool will be generic and build upon the Capabilities Approach to measure the impact on the well-being of participants and a qualitative part of the questionnaire to gain deeper insights in the experiences of the service user with the service provider, to evaluate the processes of the service provider. While evaluation will be before and after the provided service, monitoring will take place on a more frequent basis during the intervention. Monitoring will also have a quantitative part of monitoring inputs and outputs to gain insight in the progress of the SIB and a qualitative aspect to get insight in the delivered intervention. The qualitative aspect is the same as in the evaluation tool, but for monitoring it will only be used with a small sample of participants, rather than all participants as will be the case in the evaluation.

In summary, the tool will consist of different elements, which are then all developed in practical aspects of the tool. This can be summarized in table 5.1.

On top of the specifics of the tool, the governance also will look on how the evaluation is enforced. SIBs are in no way a regulated product, therefore this evaluation tool cannot be legally enforced upon the stakeholders of all SIBs. When setting up a SIB though, this design should be taken into account and clear agreements about the evaluation should be made at the start of the SIB. These agreements should make clear who the evaluator is and what the exact aspects of the evaluation tool for that SIB will be. By adding the characteristics to the legal contracts of a SIB, the individual actors can still enforce the evaluation upon the individual SIB and make sure that the evaluation that takes place is valid.

Overview of the evaluation tool of SIBs

Elements	Principles	Implementation	Contingencies
Evaluation	Evaluating Outcomes	Determine baseline data	Data availability
			Privacy regulations
			Match of characteristics with actual SIB participants
		Data collection	Validity of recorded data
		Data analysis	Expertise of analysis
			Dependencies and incentives of evaluator
	Evaluating impact on participants	Questionnaire with SIB participants	Exogenous effects on personal life of participants
			Response quality
			Response rates
	Evaluating the service provider's processes	Questionnaire with SIB participants	Dependency of participants on service provider
			Response quality
			Response rates
	Evaluating the SIB's processes	Questionnaire with SIB stakeholders	Bias of looking in hindsight based on the results of a SIB
Monitoring	Monitoring inputs and outputs	Data collection	Validity of recorded data
			Expertise of analysis
	Monitoring the service provider's processes	Questionnaire with SIB participants	Response rates
			Response quality
Governance	Stakeholders	Costs of evaluation	The overall costs of the SIB
			Available expertise of evaluators
		Stakeholder relationships	Different incentives of stakeholders
			Dependencies of stakeholders

SIB Analysis

While the first SIB was only launched in 2010, a substantial amount of attention has been given to this topic. Through public or media attention and from numerous stakeholders that have been involved in the 47 SIBs that have been launched since. In this chapter the developed evaluation tool will be tested and analysed. First it will be tested with current stakeholders of Social Impact Bonds. Second, this chapter will give an overview in the facts, when those are available. This is done through the analysis of a collected database of all current Social Impact Bonds and other resources that contain information around the individual, or in some cases combined, SIBs. Providing a case study into SIBs and how the evaluation tool would work for those SIBs.

6.1. Testing the evaluation tool on current stakeholders

Several interviews have been taken with people who are currently involved in a SIB to get a better understanding of evaluation within SIBs and how the proposed evaluation tool would work. For this, two service providers, an investor, a government and an intermediary have been interviewed, combined they have experiences with more than 15 SIB, of which three in the Netherlands. The interviewees are presented in table 6.1. Interviewees have been asked what they think is important in evaluating SIBs and their impressions and ideas around the different aspects of the designed evaluation tool. In this section the learnings from those interviews are discussed, starting with the most important aspects of evaluating SIBs and then discussing the different aspects of monitoring and evaluation. The learnings are then used to propose changes to the initial design, for the redesign of the tool in the next chapter.

The interviewees have been selected based on their different roles within SIBs so that all stakeholders of the SIB have been represented. Therefore, information has been collected from the different perspectives that exist within a SIB. Each interview has been an open interview concerning two main themes. The first theme was their impression on the most important aspects of evaluation of Social Impact Bonds. The second theme was the designed evaluation tool in this thesis and it was discussed what the opinions of the interviewees were on the different aspects of the tool. The all lasted approx-

Overview of the interviewees for this Thesis

Person	Organisation	Organisation type
Janette Powel	Social Finance UK	Intermediary
Paul Hopkins	Teens and Toddlers	Service Provider
Leo van Loon	Buzinezzclub	Service Provider
Ruben Koekoek	ABN AMRO	Investor / Intermediary
Tjalling de Vries	Gemeente Enschede	Commissioner
Christian van Esch	Deloitte	Evaluator

Top ten responses

Response (type)	# Mentioned
Monitoring is an important aspect of evaluation that should take place frequently	6
The most important aspect of evaluation is to look at the outcomes of the SIB	5
Evaluating well-being of participants would add value to the evaluation	5
Valuation of the outcomes is primarily the job of the commissioner	4
Investors should have a certain level of trust in the service providers	4
Adjusting agreements during the SIB is not possible, the agreements should stand	3
There are no incentives to evaluate the processes of the service provider	3
Evaluation should look at the actual impact a programme has	3
Softer outcomes are important for the impact of the SIB, but relatively hard to measure	3
The experiences of the stakeholders should be evaluated afterwards, look at the processes of the SIB itself	3

imately an hour and notes were taken on all interviews. Even though the number of interviewees was relatively small and have mainly knowledge on the Dutch SIBs, the different perspectives give insights in all aspects of the SIB and its evaluation.

6.1.1. Overall interview responses

Overall it is recognised that evaluation plays an important role in a Social Impact Bond, both for evaluating the results and looking back at if the SIB was executed properly, but also to provide learning lessons for the future development of SIBs [24]. Before going deeper in the individual sections of the tool and the responses of the interviewees to those specific sections, an overview is given of the ten most heard responses in table B.1.

There were obviously much more grouped responses throughout the interviews, some of which will be discussed in the remainder of this section. From the top ten responses main headlines do become clear to for the evaluation tool. First of all, monitoring is an important aspect of evaluation, but only to keep track of the outcomes. Changing agreements during the SIB is not seen as possible. The evaluation tool should also primarily focus on evaluating outcomes as that is seen as the most important aspect by the interviewees. Evaluating the well-being of participants and the effect that SIBs have on the participants is seen as important, but can't be used as an outcome indicator as it is not objective enough and there might be discussion around the methodology. Nevertheless, it can add value to the evaluation of the tool and the lessons that are being learned from the evaluation. The rest of this section will discuss the different aspects of the evaluation tool and the feedback on those aspects that have been collected through the interviews.

6.1.2. Criteria for the evaluation of Social Impact Bonds

Several important aspects of the evaluation of SIBs were mentioned, but among all stakeholders there was consensus that the primary objective should be to evaluate the outcomes and the financial returns that go along with those outcomes. In some cases this is described as the model of the SIB. It is however also recognised that evaluation brings some complexities and therefore the criteria of evaluation should be defined well.

First of all it appears to be clear that indicators should be as objective as possible not be subject to arbitrary judgement. Objective and hard indicators can be easier collected and are less prone to political influences. On softer outcomes, stakeholders are worried that especially the valuation becomes too much part of a political system that will most likely have changed by the time the SIB is in its final stages [24, 91]. Also there is little risk of bias using those indicators. There is mentioning of using softer

outcomes [36, 39, 91] as a great addition to the mix, but it is also recognised that this makes for harder measurement. An alternative would be to use them and not include them in the payment mechanism, but it would be unclear which stakeholders would then benefit of that part of the SIB and if no one does directly, it was deemed unlikely that they would pay for it. T

It is important that the main evaluation is performed by an independent evaluator. Most stakeholders discuss the aspect of trust between the stakeholders, but the independent evaluation would mainly to avoid the appearance of bias and increase validity so that the results are treated seriously. It does however depend on the part of the evaluation, as stakeholders also acknowledge that the different stakeholders have different expertises, including the evaluator [89]. For instance, collecting feedback on the processes of the intervention is mainly the expertise of the service provider and learnings of the intervention itself are seen as a role that could very well be done by the service provider. Also it would be only to a lesser extent interesting for investors and government as they would gain little benefit from such an evaluation.

The exogenous effects of a SIB are very complex and even the stakeholders wonder whether you can even find a way to completely control for this. Even to the extent that it is suggested that focusing on those exogenous effect might not be ideal, because of its complexity and because full causality will simply not be proven [24]. Usually, governments are running many other interventions that it is unclear to which effects the participants would be subject when they wouldn't participate in the SIB [89]. Therefore it will be hard to evaluate those. In one example, there is being worked with a control group and while the objective indicators can obviously be evaluated for all groups, gaining more insights from the control groups is somewhat of a hurdle because of privacy regulations [24]. Because of regulations it can also become difficult to gain certain data that might influence the model, such as ethnicity. Also, sometimes the data that is being collected by the government on the participants is not always complete [89]. The government does appear to have an interest in evaluating broader outcomes than just the set of defined outcomes. This can be split in broader outcomes that could were foreseen but were not used as an indicator for the SIB and unintended outcomes that were not foreseen. While it can be important to track the broader outcomes that are foreseen, it is harder to track outcomes that were not foreseen, because you don't know what to track. For both types of broader outcomes, the government does realise that it might be harder to track that as other stakeholders have no incentives of tracking those outcomes.

Another part of the external effects could be the collaboration with other stakeholders outside of the SIB. While the Dutch SIBs usually work directly with the participants, in the UK there was an example where the other interventions that were funded by the government and the job centres weren't always as collaborative at first [57], which adds an extra layer of complexity to the SIB.

Incentives also play a role when looking at evaluating the well-being of service users. While all stakeholders see it as an aspect that is currently missing that can add value to the evaluation process of SIBs, apart from the service provider they are also unsure how it would be added in to the process in a way that both investors and government are also willing to pay for it. Stakeholders are interested in evaluating well-being for different reasons though, investors would purely be interested to show the social aspect of the SIB. For governments it could be part of learning the unintended outcomes or tracking the foreseen broader outcomes and showing the effect a SIB as a whole has on people and society. Finally, the reason a service provider to be interested appears to be to deliver proof for the results of the intervention and gaining deeper insights what the intervention delivers. In general, using a questionnaire to evaluate the well-being before and after the intervention is seen as added value, but also something that should definitely be performed by an independent evaluator [39, 91]. However, the service provider showed concern on a single evaluation moment as the moods of participants fluctuate quite a bit and this could have a significant effect on their perception of well-being.

Costs of evaluation and therefore its practical achievability is an aspect that is definitely worrisome for the involved stakeholders. SIBs are relatively expensive products [24] and are also primarily focused on the financial returns [36, 91]. Therefore, evaluations have been focused on the aspects that are directly related to the financial returns, the indicators. In all interviews it was mentioned that additional aspects, how interesting they might be, most likely need to include an incentive for at least one of the stakeholders to include them. The same will go for including substantially more moments of monitoring or evaluation, it should be considered that this will also increase the costs. Another example of this would be when the outcomes would be dependent on different scenario's of exogenous levels (i.e. scenario's for different levels of employment). While this would be interesting, it will also increase the complexity of evaluation

as the model might be slightly different for each of the scenario's if SIBs are set up as they currently are in the Netherlands.

6.1.3. Monitoring

All stakeholders acknowledge that they are currently using a monitoring aspect in the SIBs they are involved in and that they are important for the process of a SIB. Monitoring builds up the trust amongst investors [57], gives them deeper insights [39] and help steer the SIB when needed [91]. While the UK SIB's weekly monitoring meetings have sparked a much more data driven approach from service providers [36, 57], in the Netherlands monthly monitoring would be seen as too much [91] as it distracts from what the service provider actually should be doing, giving attention to their participants.

While monitoring can provide insights and can be used to push the service provider when it looks like the targets are not being achieved, adjusting indicators is not seen as a function of monitoring by the current stakeholders. The main issue is that agreements are made beforehand and expectations are based on those agreements. When adjusting vital parts of a SIB this might affect those agreements and therefore influence the return of investors and the budgets of the government. This complexity makes that stakeholders think this cannot take place within a SIB. When it is felt that data suggests a change, this should be learned for future learnings in new SIBs though [24]. Smaller adjustments can be made however, such as stretching the eligibility criteria when it appears that people are entering the programme who technically are not eligible [89].

Stakeholders are mostly interested in the outputs and outcomes of the intervention from monitoring [39]. Other aspects appear to be less relevant for the different stakeholders. For instance, monitoring the processes of the intervention is not seen as something that adds value as evaluators and investors don't have the right expertise [89, 91], but also because by investing in an intervention this is based on the faith they have in the service provider and they should leave the provider to it [39]. Monitoring the processes of the broader SIBs doesn't make sense either as the processes can't really change while the SIB is still ongoing [24]. The aspect of trust also comes back when looking at monitoring the financial inputs and looking at the budgets of the SIB on a frequent basis. While this is something that happens regularly in SIBs in the UK, in the Netherlands it is seen as something that is less relevant. At the start of the SIB agreements have been made about the money that an intervention costs and it is then up to the service provider to deliver what was agreed upon, this is part of the trust investors put in the service provider when they make the investment [39]. Also, when they would look at the budgets frequently, that is also something that can be skewed when you would want to do harm [91].

6.1.4. Evaluation

As discussed is the evaluation of the outcomes the most important aspect of the evaluation. This should look primarily at the intended outcomes, whether they were achieved and whether they can be attributed to the intervention [39]. But it will also be important to identify outcomes past the indicators [24]. Looking back, it would be good to identify which other outcomes have been achieved, both the foreseen outcomes and unintended outcomes, positive and negative. Evaluating the effect that a SIB has on the people is seen as an interesting aspect of evaluating the impact of the SIB but also getting first insights into other outcomes that have been realised than expected. Improving well-being also might be a condition for triggering other outcomes [24, 91]. Using aspects of well-being such as health, can also mean that in the future other stakeholders might be involved such as health insurance companies, who might be interested to join when there is clear proof of them benefiting from the SIB.

The qualitative aspect of evaluation is mainly seen interesting by the government, as looking at the SIB and using that for future learnings. Evaluating the service provider's processes is again not something that is seen as adding direct value to the SIB and is more seen as the responsibility of the service provider [39]. However, for everyone the question remains whether the qualitative evaluation adds value to a SIB though and whether it would be worth the costs [24]. The main goal would be to gain better insights in the processes and use that for the development of new SIBs.

6.1.5. Conclusions from interviews

The interviews with the different stakeholders have provided some interesting insights. A lot of the stakeholders are conscious of the incentives of the involved stakeholders and that the choice of evaluation methodology but also for the aspects that are included in evaluation. Also, the costs aspect of evaluation comes back frequently as SIBs are an expensive product and therefore costs of aspects such as evaluation

should be used as efficiently as possible. The downside of the importance of costs is that stakeholders clearly see evaluation play a more important role when things would go badly and feel that the need for evaluation is less when everything is going smoothly. These insights also affect the initial design as proposed earlier in this thesis.

It can be concluded that the most important aspect remains to be the evaluation of the outcomes, the financial return will follow from that. This evaluation should be performed by an independent evaluator to avoid bias and increase validity of the evaluation. This is in line with the design as proposed. It is also recognised that those indicators can't be generic as they will be different for almost every SIB. It is however clear that those indicators should be hard and objective to allow for good measurement. These indicators will most likely be set by the commissioner. It doesn't appear to be the case that a causal model is used to construct the indicators, however the government does intend to look at broader outcomes and would like to know the unintended outcomes from a SIB. When looking at the system of the SIB, this could see two categories of intended outcomes, those that are part of the financial scheme and those that have been foreseen but not included in financial part of the SIB. Even though there is no direct financial incentive to include the latter part of the intended outcomes, as is the case for observing unintended outcomes, they are still important aspects and when the SIBs are evaluated as a whole, should clearly be included.

When looking at the different aspects and the incentives to evaluate those, it becomes clear that especially the investors won't benefit from additional parts of evaluation. The government will benefit through more insights in societal effects of interventions and learnings they can implement in the future. The service provider will be able to give better and broader insights into the performance of the intervention, which is beneficial for landing other projects. But there is close to no financial incentive in doing so for investors, especially as all parties agree that only hard and objective outcomes should be included in the financial model. Therefore, investors either have to be intrinsically socially motivated and push for additional evaluation to learn the full social impact, or an incentive should be created for them. Interestingly, these issues arise much less when talking to investors than when talking to the government, which implies that there might be different expectations from the stakeholders of each other, or at least misconceptions on their incentives.

The conclusion from the interviews for the monitoring aspect is that the interviewees expect to only focus on the outputs and outcomes of the intervention, while they are much less interested in monitoring the processes actively. This will most likely be done by asking the service provider directly, when there are questions. Therefore, monitoring should definitely take place, and every three months appears to be a reasonable frequency, but can be slightly less extensive than initially designed. This can also save costs as monitoring happens multiple times throughout a SIB and is therefore putting pressure on the total evaluation costs. Actually the interviewee representing the service provider directly thought that monitoring more than on a quarterly basis would be too much. From the main stakeholders, monitoring activities would probably also impose most time on the service provider, so increasing the frequency would also significantly increase the effort that a service provider would need to put into it. Therefore, it would make sense that a service provider would be the first to put a limit on the number of monitoring events. There is a similar issue with all stakeholders not looking at the processes of the service provider. While it doesn't add direct value to the current SIB and only additional costs, it is something that from a practical standpoint is not making a lot of sense. Therefore taking it out would be understandable. Ideally though, evaluating SIBs should not only be about optimising the process of SIBs or the value it delivers, but also look at the impact it delivers. There is probably no bigger effect on impact to be realised than to improve the services delivered to the target population. It is thus interesting that all stakeholders see no value in that aspect of evaluation. On the other side, this will also make the likelihood of this part being actively used close to nothing. Focusing on a more efficient evaluation tool that will be used in practice and delivers value to society and the SIB system will make a bigger difference from an evaluation perspective though, taking that specific aspect out of the evaluation would therefore still be defensible.

For the evaluation part of the SIB, it is clear that the initial design for the outcomes and for the well-being could remain in place based on the interviews. The qualitative aspect of the design focuses on the processes of the service provider however, while that is seen as something that doesn't necessarily have to be part of the evaluation. It is stated though that a qualitative evaluation of the processes of the SIB can be very useful though, this would therefore look at the broader system of the SIB and at the interaction between stakeholders. This would therefore shift from the participant as the unit of

observation, to using the other stakeholders of the SIB as the unit of observation.

6.2. SIB case study

While the previous section with the interviews give a qualitative view on SIBs from inside stakeholders, it is also important to look at the variety of SIBs and their goals and workings, to identify whether the evaluation tool would be applicable on all forms of SIBs. This section will look at all existing SIBs based on the social issue they are targeting and comes up with three typologies of SIBs based on that. From these typologies, the effects on the evaluation tool are discussed, which will form the final learnings for the redesign of the next chapter.

6.2.1. A high-level overview

To start off, an initial high-level overview of the existing SIBs is given in this section. The facts around the topics of investors, commissioners and service providers will be analysed further in later sections of this chapter.

In total, 47 SIBs have been launched over the world. This has happened in 12 countries, but 28 of them are in the UK. Other countries range from the US (7 SIBs) to India (1 SIB). Not for all SIBs is the data on investment known, but for the 24 SIBs where it is, there is a total of €95.13 million¹ invested, an average of €3.96 million per SIB. For 25 SIBs there is a known cap on the outcome payments that can be done in the SIB, this will mean that even if the results will exceed that cap, the commissioner will still only pay the established maximum. In total the maximum known outcome payments are €149.47 million or an average of €7.18 million per SIB.

For some SIBs there is only the investment size or the maximum of outcome payments known. In 15 cases, both the investment size and the maximum outcome payments are known. For those 15 cases the total investment was €80.86 million versus maximum outcome payments of €139.64 million. As these projects vary in duration between 3 and 10 years, the chosen method to analyse the difference is the money-multiplier, which is 1.73. It is important to note though, that these are the maximum outcome payments and not the guaranteed or expected outcome payments. Should a project not reach its contract cap because less outcomes are realised, less money will be paid by the commissioner (as is discussed in chapter 3) and therefore the actual money-multiplier is likely to not become the maximum of 1.73.

No SIB has been completely finished yet, although according to its start date and duration, the services of 16 SIBs are expected to end in 2015 of which 15 are scheduled to end this year and the Peterborough and Rikers Island SIBs will end earlier than initially planned, for different reasons. The end of the services itself does not directly mean that the results will be available as well, in some cases there is extra time needed before outcomes can be measured (i.e. a certain amount of time before an ex-offender has not been arrested). This does show that it is likely that the first results of the full duration of a SIB will become available within the next year.

Not in all SIBs could all stakeholders be identified. A total of 72 different investors are identified of which 56 investors have invested in just one SIB, six investors invested in two SIBs, four investors in three, one investor in four, two investors in five and three investors in more than 5 SIBs. The most active investor is Bridges Ventures that has done fourteen investments spread over its two investment funds, Bridges Social Entrepreneurs Fund and Bridges Social Impact Bond Fund.

Most service providers have only worked on one SIB, which is to be expected as no SIBs have been finished yet. However 4 providers have been working on two SIBs. Two of those have been involved with the both SIB tenders by the UK Department for Work and Pensions (which will be discussed later) and a third has been involved individually in a SIB but also in a partnership with two other service providers. The final service provider has been involved in two unrelated SIBs.

The third category is the commissioners. There have been 34 different commissioners, where the UK Department for Work and Pensions and UK Cabinet office together have issued fourteen SIBs in total. The UK Cabinet Office has also issued another ten SIBs from its Cabinet Office Social Outcomes Fund. It appears that only one SIB has been commissioned by a non-governmental party, the Children's Investment Fund Foundation commissioned as a private foundation a SIB in India.

Finally, in a number of SIBs an intermediary has been used to manage the ongoing processes of the

¹SIBs have been launched in countries with different currencies. For the purpose of this thesis, a fixed exchange rate is used to convert to monetised values of SIBs to Euros, the exchange rates are GBP-EUR: 1.40, DOL-EUR: 0.89.

SIBs and in some cases do performance management. In 38 SIBs an intermediary was part of the SIB, of which Social Finance UK was the intermediary for thirteen SIBs while no other intermediary worked on more than two SIBs.

Categorising SIBs

For further analysis, all available data of the SIBs will be analysed, the general overview will be collected from the overview as published online by Social Finance [70] and the data as published by the Brookings Institute [33], for individual SIBs also individual factsheets and data that could be collected will be used as input. Although the SIBs will be analysed individually, for better readability of this paper they will be grouped around SIB categories. For this the categories as defined by Social Finance will be taken as a starting point, which are Workforce Development, Housing/Homelessness, Child and Family Welfare, Criminal Justice, Education and Early Years, Health and Adults with Complex Needs [70]. Within each category, the SIBs will be processed in chronological order of the respective launch dates.

6.2.2. Workforce Development

This category contains 18 SIBs for which at least €15 million is raised in capital for SIBs where the raised capital is known. Fourteen SIBs in this category are launched in the UK, while the other four are launched in Belgium, Germany, the Netherlands and Portugal. While workforce development is a broad term, most SIBs focus on reducing unemployment either directly by training service users for a SIB (such as in the Netherlands) or indirectly by improving their educational outcomes (several SIBs in the UK). The SIBs will be analysed separately in this section and the outcomes are compared with the Capabilities Approach.

Within this category there were two major launches of first ten and later 4 SIBs by the UK Department for Work and Pensions. These SIBs will have different services that will be discussed separately but share the same outcome classifications which will therefore be discussed at once. Those two groups of SIBs will therefore be discussed within their subcategories.

The Innovation Fund – The UK – March 2012

In 2012 the Innovation Fund launched a total of ten SIBs that were designed to support disadvantaged young people aged 14 to 24 years [70]. The SIBs were split in six SIBs that were launched in round 1 in April, while the remaining four were launched in round 2 in November. The Innovation Fund was a pool of money that was committed by the UK Department for Work and Pensions (DWP) for which a tendering process was launched. The DWP provided a rate card, shown in table 6.3, that served as the list of potential outcomes that were aimed to be achieved and the maximum payment for each individual that reached these outcomes that DWP was willing to pay. These rates were also changed after the first six contracts were awarded and therefore the second four SIBs bid for different rates or even different outcomes [16]. In the tendering process, partnerships of Investors, Service Providers and in some cases Intermediaries could bid for the project based on their costs for the project, the volume and quality they could deliver, which outcomes they would focus on (a SIB could choose to focus on a specific set of outcomes) and by offering a discount on the payments per individual that achieves outcomes. The DWP awarded contracts in these outcomes based on a number of factors, including the offered discount in the bid [33]. These ten SIBs all had their specific focus, although it is not publicly available what outcomes the individual SIBs focused on or against which discount on outcomes they were delivering their services.

The more general case for the ten SIBs in the Innovation Fund pool are centred around the challenge that from 2003 to 2013 the proportion of 18-24 year olds that are NEET has risen from 10% to almost 20% [70]. The DWP wants to use interventions that target prevention from becoming NEET and even though most of the outcomes are focused on educational related outcomes, the DWP states sees this as a good outcome for the service users as it is thought to increase the likelihood that they will continue in employment, education or training [70].

Evaluating the Innovation Fund brings some complexities, but also an advantage. Although the interventions might be different for each SIB, the indicators will be constant for all SIBs in the Innovation Fund. Therefore, this will make evaluating and especially comparing the SIBs much easier than when SIBs use their own indicators. Each SIB will still see its own stakeholders, so therefore the negotiation process [41] and the outcomes of the evaluation can still differ though. It is clear that in this case the commissioner has set the valuation of the indicators, which therefore doesn't need to be part of the

Rate card Innovation Fund

Outcome	Value Round 1	Round 2
Improved behaviour at school (Measured by a letter from a teacher)	£800	£1,300
Improved attendance at school (Measured by a letter from a teacher)	£1,300	£1,400
Improved attitude to school (Measured by a letter from a teacher)	-	£700
QCF accredited entry level qualifications (below GCSE)	-	£900
Achievement of First National Qualifications Framework (NQF) Level 2 qualification	£2,200	£3,300
Achievement of First NQF Level 1 qualification	£700	£1,100
Entry into first employment including a training element	£2,600	£3,500
Entry into sustained employment	£1,000	£2,000
Completion of first NQF Level 3 training/vocational qualifications	£3,300	£5,100
Successful completion of an ESOL course	£1,200	-
Entry into education at NQF level 4	£2,000	-

general evaluation [90]. One of the main issues for the SIBs in this group is that the end goal, reducing the likelihood of becoming NEET, can only be tested by the fact whether the young people become NEET before their 25th birthday, in some cases this will be ten years after the end of the intervention. As investors are being repaid based on indirect indicators towards this end goal, this will shift the risk of exogenous effects towards the government. It could be argued that this should lead to either outcomes that are measured at age 25 [28], or a potential refinement of the end goal of the SIBs to, for instance, the improvement of secondary education outcomes, to enable appropriate evaluation.

Indicators range from clear and hard outcomes, such as the improvement of attendance, the achievement of educational results and finding employment, to softer outcomes such as the improvement of behaviour and attitude. Especially the latter two are subject to subjectivity as they are measured by a letter from the teacher, which is collected directly by the service provider. This creates a risk for bias in the results and therefore should be treated very carefully in the evaluation. The evaluation is executed by the commissioner, but data is collected by the service providers of the SIB. As most indicators are objective, this should not be an issue concerning objectivity, aside from the two previously mentioned indicators. It does not use a third party evaluator as stated in the criteria. For proper evaluation, it might have been wise to either use a third party evaluator, or only use the objective indicators, to avoid the potential for bias in the data. Although detailed data is not available for all SIBs, at least some of them also use monthly monitoring in which all stakeholders are involved to discuss the inputs and the outputs of the project and the learnings that should be drawn from that for future execution of the project.

There might be an issue with establishing causality and actual value created by this model of SIBs. Outcomes are paid based on each achieved outcome, and the valuation of these outcomes include the determination of dead weight (i.e. how many outcomes would be achieved without the use of an intervention). This is calculated based on historical baseline data [16]. It is unclear though, how these numbers are exactly calculated and therefore the added value of the interventions cannot be evaluated appropriately. Also, this increases risks such as cherry picking of participants and the influences of exogenous factors on the outcomes of the intervention. Therefore it might be concluded that although the SIBs in the Innovation Fund appear to go through the right steps in the evaluation process and the costs associated with these apparently fall within the budget, very little is done to reduce the potential of biases in the outcomes of the evaluation.

Innovation Fund and Capabilities As there is no clear distinction in the outcomes that each individual SIB from the Innovation Fund targets, the outcomes as presented in the outcomes rate card can be compared to capabilities as a whole. First of all there are a couple of outcomes that are fully

focused on educational attainment. This clearly links back to the capability of Senses, Imagination and Thought [12, 58, 61]. Improvement of behaviour and attendance can be signs of someone's aspiration and entry into employment can both be linked affiliation [56, 61]. When it is clear what an individual SIB focuses on, this will be mentioned in its individual discussion.

Career Connect – Greater Merseyside, UK – April 2012 A group of 8 investors provided the upfront capital of £1.5 million for the SIB in which services are provided by Career Connect. It will also use a intermediary in Triodos Bank UK [33]. In total it aimed to service 3,900 14-19 year olds who are NEET or identified as a risk of becoming NEET in the Greater Merseyside area, with a focus on ex-offenders, those in or leaving care, and those with learning difficulties [70].

While running the intervention, it became clear that a focus on the 14-16 year olds was the most effective approach and therefore the focus of the intervention was moved to this specific group. The service tried to work with the young people to re-engage them with education and increase their school outcomes. This should also result in savings for the government through reduced welfare payments over time [14]. Financially the programme delivered good results as it was able to repay all provided capital to investors plus an unknown percentage of interest, directly after finalising the service delivery.

Nottinghamshire Futures – Nottingham, UK – April 2012 Over 3,000 young people aged 16 to 24 who are no longer in school are targeted by Nottinghamshire Futures, which makes it one of the larger SIBs to date. A total of £1.7m was raised for this SIB [70]. It also has a fairly unique structure as it has a single investor that is not a foundation or private investor, but the programme is pre-funded by the Nottingham City Council [33]. It is therefore more a direct contract between national and local government based on a pay-for-performance structure. This begs the question what the added value of a SIB funding was from a government perspective as it merely shifts the costs from local to national government. As discussed earlier, another advantage may arise from better performance management of the services, through more strict reporting requirements towards DWP.

During the programme, it was identified that there weren't enough young people in the previously identified target group and it was therefore decided to shift to target school age young people and focus on educational outcomes instead. Eventually the intervention has helped 1,962 young people and generated £2.85m in outcome payments [46]. It is unclear how these outcome payments compare to the actual costs of the programme. The intervention itself appears to focus on preparing young people for work or training [33] and therefore it could be argued that the main goal is employment and no educational outcomes. In this case, the main relation to the capabilities approach will be that with the capability Affiliation. It is unclear however to what extent the services changed towards education after its renewed focus on school-aged young people.

This SIB is also one of the few SIBs that has published some preliminary results of the work. It is unclear what the outcomes are for the specific 1,962 young people that participated in the service. Overall the programme helped the Nottingham City Council to get better understanding of their young people in their council as they did not know the employment or education status of 6.6% of the 16-18 year olds in 2011, those had fallen off the radar. In 2014 this had dropped to 2.2%. Over the same time period, they also saw a reduction of NEETs by 31.6% [46].

BEST Network – West Midlands, UK – April 2012 Another large SIB is that of BEST Network. It targets to help 2,897 disadvantaged young people and raised £3 million in capital. Interestingly it also proposes a rather low investor return with maximum outcome payments only being £3.4m, this could maximally lead to a return of c.4% IRR [70]. The intervention will eventually provide apprenticeships and employment through a programme called the Advance Programme. Schools or local organisations refer young people to the programme, but the programme also focuses on school attendance and performance as a prerequisite for the Advance Programme and to promote participation in apprenticeships or employment afterwards [33].

An interesting aspect to this SIB is that services are actually provided by a network of service providers that are all part of the Birmingham Employment, Skills and Training Network (BEST Network). This obviously adds to the complexity of managing stakeholders, which in this case done by APM UK Ltd., which is a prime contractor of service providers and therefore is already linked in with the service providers. Nonetheless, this is a challenging task for performance management.

By targeting school attainment and following up with apprenticeships and employment programmes, the SIB aims to achieve both improvement in education and employment. Looking back at the different programmes, it is clear that the attainment of secondary education is a clear example of the capability Senses, Imagination and Thought [12, 47, 58, 61]. Also employment is seen as a proxy for the capability of Affiliation [61].

Tomorrow's People – Shoreditch, UK – April 2012 In one of the most up and coming area's of London, Shoreditch, Tomorrow's People targets to work with 1,050 14-18 year olds who are identified by 11 participating school as low-attaining and most at risk of dropping out or even becoming unemployed. For this, £0.9m is invested [70]. Tomorrow's People will place coaches in the schools where they become the main point of contact for the at risk individuals and provide personal guidance. Through guidance, improvement in positive thinking and more engagement with school, 60% of the target group achieved five A*-C GCSEs against a target of 30% [70].

Even though the SIBs issued through the Innovation Fund focus on employment and workforce development, it is clear that through the target and success rates of secondary school attainment, the Tomorrow's People SIB has been focusing mainly on education. Secondary school attainment is part of the Senses, Imagination and Thought capability [12, 47, 58, 61] and appears to be the main focus of this SIB.

Community Links – East London, UK – April 2012 Both in terms of investment size as in the size of the target group, Community Links is a relatively small SIB within the Innovation Fund. With the investment of £0.37m, 740 young people aged 14 to 19 will be offered an intervention though. The target group is being referred through schools or other agencies as at risk of becoming NEET, but young people that have already left school can also refer themselves to the programme [70].

The programme centres around supporting young people in the transition between school and work. It is unclear however which outcomes are exactly targeted by the programme. In line with the general outcomes fund they will be either targeting educational attainment or employment, these outcomes are identified to be in line with the Senses, Imagination and Thought, and Affiliation capabilities from the list of central capabilities.

Perth YMCA – Perthshire & Kinross, UK – April 2012 The SIB with the Perth YMCA is small, supporting 300 young people, but involves the bigger YMCA (Scotland) which in itself is a significantly larger organisation that support disadvantaged individuals. The investment size is unclear, but twelve investors all contributed between £5,000 and £50,000. The programme works with young people aged 14-17 that are disengaged high-school students and young people aged 16-24 who are currently NEET. Although it says to focus on both educational and employment, the first results show that in the first six months, 61% of the first 86 participants were placed in employment and 100% of them had retained employment [79].

This SIB in itself brings some unique aspects with it as well. It shows that a small SIB can work and can work with a large group of (very) small investors. Secondly, it shows that SIBs can potentially bring new sources of capital into existing organisations such as the YMCA, as most of the investors in the SIB have not been donors to YMCA previously [79].

Adviza – Thames Valley, UK – November 2012 Adviza is the service provider for one of the SIBs that were part of the second round of the Innovation Fund. Investors funded the programme for £0.8m to help 1,500 young people. Adviza offers teenagers of 14-16 years old that are at risk of becoming NEET and has the aim of building resilience, confidence and aspiration. The goals are very similar to that of the Teens and Toddlers programme, but the approach is different. Adviza hopes to reach their goals through ongoing support, work experience and mentoring opportunities [70].

As there is a link between young people with low levels of aspiration and the risk of them becoming NEET [56], aiming to increase the levels of aspiration seems a sensible approach towards employment outcomes, which in itself is linked to the Affiliation capability [61]. It also appears to be a method that aims on preparing the young people for employment rather than putting them in actual employment, which is more in line with the capabilities approach itself.

Teens and Toddlers – Greater Manchester, UK – November 2012 The Teens and Toddlers SIB raised an initial £800K to support 1,100 young people aged 14-15 that are identified as under-performing at school or have other issues and are therefore at risk of becoming NEET [70]. During the intervention the teens will get the responsibility of overseeing a toddler on a weekly basis in a nearby nursery. At the same time they will receive support on topics such as Maths and English [98] and weekly sessions centred around building their self-esteem and decision making skills are provided [36].

The focus of the programme is on improving education rather than providing direct workplaces. By increasing decision making skills but also by the focus on educational skills (Maths and English), this SIB is mostly aligned with the Senses, Imagination and Thought capability [12, 47, 58, 61]. However, even though it doesn't appear to focus on employment, the programme is centred around building self-esteem and this is recognised by Nussbaum as the second part of the Affiliation capability [47].

3SC Capitalise – Cardiff & Newport, UK – November 2012 The service provider Dyslexia Action & Include, is set to help 720 14-16 year olds that are at risk of not gaining a GCSE in English or with literacy or self-esteem issues [70]. The intervention tries to improve literacy and academic achievement through the services of specialist teachers, for instance for young people with dyslexia. In total £420K was invested in the SIB. The targeted outcomes are all focused on education, eventually working towards entry level or level 1 and 2 qualifications [1]. Literacy is seen as one of the primary predictors for future NEET status of an individual and is therefore an important aspect to focus services on. Nonetheless, literacy is seen as part of education and this SIB therefore directly works on improving the capability Senses, Imagination and Thought for the participating young people. Indirectly, improving literacy will also affect the Affiliation capability as improved literacy will improve the likelihood that an individual is able to secure employment in the future [15].

Prevista – West London, UK – November 2012 Similar to the BEST Network SIB, the Prevista SIB works with different service providers that deliver interventions. In total six partners will run a set of diverse interventions that range from sport to graffiti art projects [70]. Overall, the programme focuses amongst others on workshops on gang activity, further education and work integration programmes. Although the latter appears surprising as the programme works with 14 and 15 year olds [25]. Participants are referred by schools and are identified by having issues such as offending history and substance misuse [70]. Seeing the age of the participants, it is most likely that the main focus will be on the outcomes that are in line with secondary education attainment and with the Senses, Imagination and Thought capability. However, it is unclear yet whether that is actually the case.

Juvat – Augsburg, Germany – September 2013

100 Unemployed young people that are under the age of 25 years old and also have had no contact with employment agencies of any kind, are targeted by the SIB in Augsburg. With an investment of €300K this group will be supported through intense support to find an apprenticeship or employment and continued support once such a placement is found [33].

The SIB aims to place at least 20 service users into work or apprenticeship for more than 9 months during the timeline of the project [83], but will try to also execute qualitative evaluation, although it is unclear how this will be used or how outcomes will be measured [33]. Since there are no other outcomes apparent, this is the only visible capability. Although the service seems to be in line with increasing people's choices, the single outcome could potentially increase the risk of somewhat forced employment for the service users to realise the desired financial outcomes. It is unclear how the Juvat SIB goes about that.

Evaluation appears to take place both quantitatively through hard indicators (sustained employment or apprenticeships) as well as qualitatively. Questions remain around the qualitative evaluation, which is important as on smaller groups such as in this SIB, the qualitative evaluation can provide crucial insights but should be designed well as wording and methodology of qualitative evaluation can influence responses [2]. The evaluation will be performed by an external evaluator [33] and will be based on administrative data, so the SIB will work with baseline data rather than a control or comparison group. As this is a relatively small SIB, but still a third party could be hired for the evaluation, this shows that in terms of costs, even with a third party, evaluation should not necessarily be an issue.

There is only one repayment moment, after the 28 month contract. This will likely be the moment of the final evaluation as well. It is unclear however, if there are other evaluation moments, also there

is no mention of more frequent monitoring or negotiation on the outcomes of the evaluation. With a contract of 28 months, more frequent monitoring and when needed adjustment of the intervention will be highly important and therefore should definitely take place.

Buzinezzclub – Rotterdam, Netherlands – December 2013

In Rotterdam, c.€700K was invested to reduce youth unemployment in the City of Rotterdam. This intervention provides a training programme for youth between 18-27 years old and who are unemployed for 0-24 months. During the programme, 160 young people will be going through a training where they are helped to find their passions and main interests and ideally start a business in that. In many cases starting a business is not a viable option however and with the individual it is decided what he/she wants or is in their best interest for the future and the programme will focus on going into education or employment [93]. Interesting is that the investors have provided only 60% of the capital while the service provider is responsible for the remaining 40% of the needed working capital [70]. Therefore the service provider is taking up some of the risk that is usually transferred to the investors. If the service provider is treated as a regular enterprise, taking risks is something that is normal and sometimes even expected from enterprises. For charities however, this can be seen different because their role in society, but mainly because they are not able to save enough upfront capital to do that investment.

There is a clear focus on employment and therefore with the capability of Affiliation [61]. It also shows that the approach can be seen somewhat in line with the Capabilities Approach, because employment is not forced upon the unemployed even though the main focus is to reduce unemployment. The focus is around improving decision making and giving service users the right toolset for their future. The outcomes however purely steer on the duration of unemployment and it should therefore be ensured that to improve outcomes for the SIB itself, the service user will remain central and employment will not be arranged against the will of a participant.

Evaluation is based on historical data. In this specific case, baseline data is calculated based on the personal characteristics of an individual [93]. This comes closer to the characteristics of a control group, as the population are used, but the characteristics are based on individuals and not the whole population, practically reducing the size of the control group to 1. The evaluation is performed by a third party organisation, Deloitte, and monitoring on the inputs, outputs and outcomes take place every three months. The SIB is split in two groups of 80 young people and they are evaluated separately, just as there are two payment moments. Because the historical data is not updated, this leads to four evaluation moments after the determination of the baseline data, before and after the two groups.

Duo for a Job – Brussels, Belgium – April 2014

The Duo for A Job SIB was a very small one and only raised €234K in upfront capital. It still stands to help 180 unemployed 18-30 year old migrants from non-EU, US or Canadian nationals [70]. Outcomes are measured by comparing the results with a matched comparison group to prevent cherry picking in the selection of the 180 service users. The service provider matches the service users with retirees that worked in the sector of the user's interest and uses them as advisor for the service users for 6 months. After that, the service provider will follow-up on an individual level to support the user's job-searching activities [33].

The outcomes are directly measured on employment and is defined as having more than 90 days of employment in the year following the programme [33]. The service is fully focused on creating employment for unemployed and this is in line with the Affiliation capability [61]. Since there are no other outcomes apparent, this is the only visible capability. Although the service seems to be in line with increasing people's choices, the single outcome could potentially increase the risk of somewhat forced employment for the service users to realise the desired financial outcomes. It is unclear how the Duo for a Job SIB goes about that.

In this SIB, a quasi-experiment is being executed with a matched comparison group. This comparison groups allows for more valid evaluation, while the relatively small number of participants makes it easier to use a comparison group. The evaluation is performed by an external organisation, the Brussels Observatory of Employment, which is still linked to the commissioner and therefore the outcomes are verified by an independent validator. This combination has the potential of removing bias from the evaluation as much as possible. The outcomes can also be evaluated directly as the only outcome that is being measured is that of days of employment in the year following the programme. As the focus is on the year after the programme, this also allows for direct measurement of employment rather than

Rate card Youth Engagement Fund

Outcome	Value (change vs. Innovation Fund)
Improved behaviour at school (Measured by a letter from a teacher)	£1,100 (-£200)
Improved attendance at school (Measured by a letter from a teacher)	£1,400 (+£0)
Improved attitude to school (Measured by a letter from a teacher)	£400 (-£300)
QCF accredited entry level qualifications (below GCSE)	£1,000 (+£100)
Achievement of First National Qualifications Framework (NQF) Level 2 qualification	£3,900 (+£600)
Achievement of First NQF Level 1 qualification	£1,500 (+£400)
Entry into first employment including a training element	£3,400 (-£100)
Entry into sustained employment	£1,900 (-£100)
Completion of first NQF Level 3 training/vocational qualifications	£2,700 (-£2,400)

indirect measurement as with the Innovation Fund SIBs. It does not appear to be the case that processes are evaluated though, or a larger causal model is established. This could lead to a large number of unintended outcomes that are not accounted for but will take place. Examples of that would be lower well-being due to forced employment or young people moving to education because they learn that they are not ready for a job.

The Youth Engagement Fund – The UK – March 2015

The Youth Engagement Fund can be seen a follow up fund by the UK DWP to build upon the knowledge that was gathered through the first ten SIBs in the Innovation Fund. The tendering process and the chosen outcomes were exactly similar to the Innovation Fund, but there was a change in the proposed tariffs for the different outcomes and the new maximum payments per outcome are stated in table 6.4. The changes in valuation were based on feedback from the experiences in the Innovation Fund on the achievability, but also relevance of the different outcomes for the target groups.

There are some variations in the pricing, but more interestingly, three out of the four SIBs that were awarded a contract are with service providers that also were awarded in the SIB the time before. Although there are no available explanations for this, there are two types of learning that could potentially be taking place. First, awarding a new contract to practically the same SIB gives the opportunity to see whether the involved stakeholders have learned from the first contract and are able to improve their results based on those experiences. Secondly, as bidders also include a discount, it could be possible that existing SIBs were able to offer a better discount on the proposed maximum tariff, because they have learned from the first SIB that they are able to deliver their services and achieve outcomes for a significantly cheaper price than the maximum tariff.

The evaluation system is completely similar to that of the Innovation Fund as the commissioner is also exactly the same. There is a shift though in the outcomes that can be achieved. First of all the valuation of some of the outcomes have changed. As discussed in the second chapter of this thesis, it can be argued that setting the valuation is fully up to the commissioner and it might be that learnings from the Innovation Fund have led to the changes, but the Innovation was not finished when the Youth Engagement Fund was launched so the final evaluation of that has not taken place yet. It is somewhat unclear how the new rate card was established, but if that was indeed the consequence of early learnings of the Innovation Fund, that shows that the monitoring function has been used appropriately and learnings are being implemented. If that is not the case, it can be seen as an effect of the evolving market, but more transparency might be expected on that evolution to make sure that the market as a whole can make use of those developments and the rationale behind it.

Career Connect – Greater Merseyside, UK – March 2015 As the YEF SIBs have just started, we only know the size difference with the same programme in the Innovation Fund round. The new Career Connect SIB will target to help 4,000 young people versus the 3,900 that they targeted in the same time period the previous time around. This means that the capacity of the organisation for the SIB has grown by 2.5% since three years earlier. It is unclear if there are also changes in investment size, but it is assumed that the general programme is the same as before.

Teens and Toddlers – Greater Manchester, UK – March 2015 Teens and Toddlers appears to have scaled up significantly from the Innovation Fund. It is now aiming to support 1,600 disadvantaged young people [70]. This is an increase by 45% from the 1,100 that Teens and Toddlers supported in the first time around. This could suggest that the capacity building that Teens and Toddlers experienced through the involvement from the different stakeholders [36], did pay off and enabled the service provider to deliver the programme on a larger scale than before being involved in SIBs. Such growth also brings in risks, as the CFO Paul Hopkins agrees, because by the end of the YEF SIB, Teens and Toddlers will have had SIBs that make up significant parts of the organisation's total income for over six years. This also means that it will need to seek for either new large SIBs or specific contracts to follow up the income streams from the two successive SIBs and keep the organisation in tact as it is.

Futureshapers Sheffield – Sheffield, UK – March 2015 The Futureshapers SIB is the only SIB that was not involved in the Innovation Fund. The service provider Sheffield Futures sets out to help 1,300 and has at least an investment by Key Fund of £225K. It works with two groups. First it will look to improve school behaviour, attendance and attitude, and the achievement of secondary educational qualifications for 14 and 15 year olds. Secondly it will also assist 16 and 17 year olds to achieve vocational training and entry into employment [38]. By splitting the service clearly in these two groups it will target improvement in the capability of Senses, Imagination and Thought for 14-15 year olds, while it will focus mainly on improving the Affiliation capability for 16-17 year olds.

Prevista – London, UK – March 2015 Although the size of the Prevista SIB in the Innovation Fund is unknown, the SIB size was described as large, while for instance that of Teens and Toddlers was described as medium [25]. As the Teens and Toddlers supported 1,100 participants, this implies that the Prevista SIB supported more than 1,100 people. The Prevista SIB in the Youth Engagement fund targets to support 1,000 [70] people and is therefore likely to have become smaller than before. Because of a lack of information this is however not confirmed.

Conclusions Workforce Development

Divided over a total of 17 SIBs in the category workforce development, at least 19,547 disadvantaged young people are being supported and if the Prevista SIB from the Innovation Fund round can be included, this will likely go over 20,000. As the category and the large involvement of the UK Department for Work and Pension might suggest, most SIBs target a reduction in unemployment either directly or indirectly, which can be linked with the Affiliation capability. Especially in the UK however, the outcomes that are defined also steer towards an improvement of secondary school attainment and therefore more towards the Senses, Imagination and Thought. Improving literacy and the attainment of 5 A*-C GCSEs are however seen as important predictors of a person becoming NEET in the future [15] and therefore improving those aspects will have an indirect effect on future employment and therefore the Affiliation capability.

While not all SIBs have published their exact services and targeted outcomes, it appears as if 8 SIBs are directly targeting outcomes that are in line with the Affiliation, while 9 SIBs target outcomes that are directly in line with Affiliation but are in itself indirectly improving the Affiliation SIB as well ². It is interesting that due to the structure of the SIBs that are issued by the UK DWP, more SIBs focus directly on Senses, Imagination and Thought rather than Affiliation, which might be expected beforehand.

As most SIBs in this category are part of the Innovation and Youth Engagement Funds, they also share the same evaluation methods which is based on baseline data without comparison groups.

²As some SIBs target different types of outcomes and for some SIBs outcomes are unknown, the number of outcomes might not be in line with the total number of SIBs

Evaluation in those 14 SIBs is done directly by stakeholders involved in the SIB and data collection is often done by the service provider itself. This increases the risk of bias and therefore lowers validity of the evaluation. The best evaluation example is that of the Duo for a Job SIB in Belgium which uses a matched comparison group and a third party to verify the results. The other two SIBs in this category do make use of a third party evaluator, which is positive. They also use baseline data to compare outcomes with, which can be seen as sensible to reduce costs and increase the practical achievability of evaluation and scalability of the SIB. In only one case does there appear to be a qualitative evaluation, while this would be an important aspect of good evaluation. Overall it can be concluded that based on the available data, there are definitely good aspects to the existing evaluation methods, but to reduce potential bias and increase validity, big steps forward can still be made.

6.2.3. Education and Early Years

Some of the workforce development SIBs are focusing on education, but the main aim when they were being developed, was that they would have at least a long-term outlook on improving Employment. The SIBs in the category education and early years, mainly focus on education but the service users are also younger. SIBs focus on working with children in or even before primary education to improve educational outcomes and the children's opportunities for their future life. In total there are four SIBs that fall into this category, although they are very different in size, with two of the smallest and two of the largest SIBs in terms of investment size.

Junior Code Academy – Lisbon, Portugal – January 2015

The smallest SIB to date is that of the Junior Code Academy in Lisbon. As a pilot SIB it raised €120K to support 65 primary school students aged 8-9, but with the goal to scale the programme to 90 schools in Lisbon by 2017 [70]. Through a 30-week computer programming intervention the primary school students will learn programming skills [33]. The goal is to improve the cognitive skills and school performance and retention of the service users.

Outcomes are measured based on the logical thinking and problem solving skills and the national exams in Portuguese and Mathematics, compulsory for all 4th grade students in Portugal, which take place two terms after the intervention is finished [33]. The outcomes are clearly related to the Senses, Imagination and Thought capability and they are measured based on literacy (the Portuguese improvement) and mathematics training, which are identified by Nussbaum as central in this capability [47]. Although mostly it is focused on secondary school attainment, Sen argues that school attainment for children should definitely be seen as an improvement of their capabilities and therefore primary school cannot be excluded in that. Therefore, the SIB directly aims to improve the capability of Senses, Imagination of Thought for the participants. Improving literacy also improves the chance that people are able to secure employment later on in line and the improvement of literacy could therefore also be seen as indirectly improving the Affiliation capability.

This SIB is a clear example of evaluation where it is chosen to only evaluate specific items of the causal model. While there might be a clear link between better developed cognitive skills and school performance/retention, the latter two outcomes are not evaluated and are further down the line. All potential indicators are objective and measurable, but the evaluation focuses on the improvements in Portuguese and Mathematics. The advantage of that is that it will be easier to look at causality, as the outcomes can be measured directly after the interventions with less exogenous risks. Improved school performance and school retention will have effect further down the line and are therefore definitely more prone to exogenous risks and harder to measure. The evaluation itself uses a randomised control trial though, which is rarely seen in SIB but highly increases the validity of the outcomes. Also, the evaluation is performed by a third party organisation, Nova School of Business and Economics [33]. As with the SIB in Belgium, it is clear that this SIB targets a smaller number of people and it would be very interesting to see when the SIB does reach its intended scale of 90 schools, whether it will still use a randomised control trial for evaluation. For now, it exceeds the criteria as set in this thesis, which can be seen as a good thing.

Girl's Education – Rajasthan, India – May 2015

The SIB in Rajasthan was the first so called Development Impact Bond, a SIB that is focused on developing countries, that was officially launched. Although the first announcement about it was made in June 2014, service delivery started in May 2015. Although the upfront investment of \$238K is

relatively low, the service provider Educate Girls intends to help girls in Rajasthan keep in school and improve learning outcomes, targeting to retain 10,000 girls at school and improve basic English, Hindi and math for 20,000 boys and girls in 150 poorly performing local schools [53]. If successful, the investor can see a return of 7-13% paid by the outcome payer Children's Investment Fund Foundation. Apart from working in Development Impact Bonds, this outcome payer is very interesting as well, as it is not a governmental organisation but a NGO that aims to improve the lives of children living in poverty in developing countries [86].

By improving educational outcomes in English, Hindi and Math, these outcomes would be in line with improving education and therefore with the capability Senses, Imagination and Thought. However, by focusing on solving the high percentage of girls that drop out of school, it is also focusing on improving gender equality in India. When looking at the description of the Affiliation capability by Nussbaum, it reads *...This entails provisions of non-discrimination on the basis of race, sex, sexual orientation, ethnicity, taste, religion, national origin and species* [47]. Therefore it can be argued that if this SIB improves gender equality, it might also improve the capability of Affiliation for the girls that are participating. Improving literacy also improves the chance that people are able to secure employment later on in line and the improvement of literacy could therefore also be seen as indirectly improving the Affiliation capability.

Even though there is mention of an independent evaluator [86], it is not clear who that is or how evaluation will be performed. Outcomes can be measured objectively as basic levels can be tested, as long as the tests are standardized and don't change over time. Seeing the scale of the intervention with up to 20,000 participants and 150 local schools involved, it appears unlikely that the evaluation will use a comparison group. Nonetheless, it is unclear how the evaluation exactly takes place. There is use of an advisory board that frequently meets, so there is some form of monitoring, but methodologies or stakeholders involved are unclear.

Chicago Public Schools - Chicago, USA – October 2014

So far, the Chicago Public Schools SIB has raised the most upfront investment of all launched SIBs, with \$16.9 million raised. In total it aims to support 2,600 families with young children in Chicago [70]. The intervention aims to implement high-quality pre-school education for four year old children coming from families living below 185% of the federal poverty level. Outcomes that are targeted by the SIB are threefold. First it will be aimed to increase the kindergarten readiness of the children, secondly an increase of third grade literacy (measured by a third grade reading test) will be strived for and finally, a decrease in the use of special education services by the target group will be measured [18]. Although the outcomes of increased literacy and reduced use of special education services appear to be outcomes, it seems that the first outcome, increasing kindergarten readiness, is more an output than outcome as it is unclear how that is measured other than completing the offered programme in the SIB. The outcomes do steer towards improving the educational capabilities of the participants with the aim of improving future school attainment. Therefore, the outcomes appear to be in the line of the Senses, Imagination and Thought capability.

When there is an apparent mix between outputs and outcomes, it is even more important that evaluation and data collection takes place properly to make sure that the right conclusions can be drawn from the project. There is no evaluation method publicised though and therefore it is not only unknown how evaluation takes place, but also when and by whom. For a product (SIBs) that is in such early stages, transparency on all aspects is vital for the development of SIBs and implementing learnings. The Chicago SIB lacks this and therefore no comments on the evaluation method can be made.

Utah School District YMCA & Lit'l Scholars – Salt Lake City, USA – September 2013

Another educational SIB in the USA, in Salt Lake City. \$7 million was raised to support 3,500 3-4 year olds that come from low-income families and are predicted to need special education or remedial education in the future. It was shown that in the Granite School District 33% of students from low-income students were likely to make use of special education services, while 95% of the service provider United Way of Salt Lake no longer needed those services, saving over \$31K per child in 12 years [88]. It tracks the participants from preschool through 6th Grade to see if they make use of special educational services and outcomes are measured for reduction in the utilization of those services. However, the outcomes are not directly measured on educational attainment, they are designed to improve the likelihood that

a child is able to attain educational qualifications. Therefore, achieving those outcomes could still be seen as in line with the Senses, Imagination and Thought capability.

Outcomes are evaluated through administrative data that serves as a matched comparison. This data is validated by the Utah State University as an independent evaluator [33]. By tracking the use of special educational services all the way through 6th grade, evaluation allows to measure direct effect rather than just indirect (decreasing the likelihood of using special educational services), which is a good aspect of this evaluation method. The matched comparison approach makes use of historical data, to improve overall validity it should also include overall trends during the time the intervention takes place, such as national trends on the use of special educational needs, to decrease exogenous risks such as that of generational differences.

Conclusions Education and Early Years

Education has a clear link with the capability Senses, Imagination and Thought, but in its operationalisation this appears to mostly be measured through secondary school education, which can be more found in outcomes of the workforce development SIBs. When looked at the definition by Nussbaum on this capability however, it appears that other initiatives to improve the educational capabilities can still be seen as improving that capability. All four SIBs in the education and early years SIB seem to strive to achieve outcomes that are in line with the Senses, Imagination and Thought capability. Three of the four SIBs (all, except from the SIB in Salt Lake City) also specifically aim to improve literacy for children. As literacy is an important predictor for future NEET status and is best improved at early years [15], successful outcomes for these SIBs should also improve the possibility of the participants of finding employment and therefore contribute to the Affiliation capability. Although there is no reference to the Capabilities Approach in any of the SIBs whatsoever, it is clear that the importance of the link between education and literacy is understood by the SIBs in the education and early years category, in light of improving the capabilities of the participants this also appears to be a wise decision.

Of the four SIBs in the Education and Early Years category, for two it is somewhat unknown how the evaluation takes place, which is especially worrisome from the point of view of the formative function of evaluation. While evaluation for those SIBs might take place appropriately and the formative function is used properly within the SIB, for the development of SIBs as a whole it is important that the system also learns from the outcomes of the different SIBs and the methods they have used to get there. The only example of a SIB using randomised control trials is used although this SIB has a relatively small target population, but also the Utah SIB that uses matched comparison to historical data appears to have set up the measurement of outcomes and the evaluation of that appropriately.

6.2.4. Housing/Homelessness

Ten SIBs have been launched in the housing/homelessness SIBs. With a total of over €13 million in capital raised, the number of people that are receiving services are just over 2500 [70]. This represents relatively high costs per service user, however, housing problems especially in the UK can be expensive for the government and expensive to solve as well. With nine out of ten SIBs in this category taking place in the UK, the costs are therefore not necessarily surprising.

As with the workforce development SIBs, the UK government has developed a fund that commissioned in this case seven SIBs, the Fair Chance Fund. It was designed to award 5-15 applications of £500K to £3m each. [17].

Fair Chance Fund

The commissioned SIBs in the Fair Chance Fund all will target homelessness for 18-24 year olds, the total estimated population in that group is estimated on 60,000 young people. The process of commissioning SIBs in the Fair Chance Fund is very similar to the Innovation Fund and also established a rate card of tariffs against which bids could be placed in the tendering process. The Fair Chance Fund obviously has different targets, the rate card that was established for the Fair Chance is as shown in table 6.5 [17].

Not only does this rate card have significant more possibilities than the rate card used in the Innovation Fund or Youth Engagement Fund, there are two other important differences to be noted. First of all a choice has been made to split most outcomes to also look for sustainment of employment, volunteering experiences and accommodation. Rather than just choosing one cut-off, SIBs will be rewarded for more durable solutions with clients. The second important difference is that even though

Rate card of Fair Chance Fund

Outcome	Value
Initial Assessment Fee	£500
Second Assessment	£500
Third Assessment	£200
Move into accommodation	£500
Accommodation sustained for 3 months	£1,500
Accommodation sustained for 6 months	£1,500
Accommodation sustained for 12 months	£1,500
Accommodation sustained for 18 months	£1,500
Entry into Education or Training	£500
Individuals first Entry level qualification (inc. Maths and/or English)	£1,500
Level 1 Qualification (e.g. NVQ)	£2,500
Individuals first full level 2 or equivalent	£3,500
6 weeks volunteering (6-16 hours)	£500
13 weeks volunteering (6-16 hours)	£500
20 weeks volunteering (6-16 hours)	£250
26 weeks volunteering (6-16 hours)	£250
Entry into employment	£500
13 weeks P/T Employment	£3,000
26 weeks P/T Employment	£2,000
13 weeks F/T Employment	£4,500
26 weeks F/T Employment	£3,500

a SIB is based on performance by results rather than pay for service delivery alone, the first set of outcomes appear to function as a fee for delivery of the service, by paying for just doing assessments. It is not really clear how those payment fees are linked to the outcomes for service user and what their function is.

For individual SIBs also this time around, there is little known about which outcomes the SIBs target. When looking at the capabilities approach however, the overall capabilities that can be linked to the rate card can still be established. The main goal for the SIBs in this fund is to reduce homelessness for 18-24 year old and provide sustained accommodation for the service users. When looking at the capabilities and the definition by Nussbaum, having adequate shelter is part of the Bodily Health capability [47], however none of the operationalisations have used that as a measurement so far. Furthermore, there is a second set of outcomes that is centred around educational attainment, which can be linked to the Senses, Imagination and Thought capability, as can be seen before. Finally, there are outcomes around employment and volunteering, which has clear links with the Affiliation capability.

As the Fair Chance Fund was launched in between the Innovation Fund and the Youth Engagement Fund, the structure is also very similar to those two SIB Funds. Obviously, the target population and the social issue are both different and therefore the outcomes are very different. From an evaluation aspect the most important difference is that the outcomes in the Fair Chance Fund are measured by sustainment, which also changes the way evaluation should be performed. The proposed method in this thesis of evaluation before and after the intervention is can only take place when the evaluation after the intervention takes place 18 months after the intervention as that is the longest delay an outcome can take place. This however is not ideal for investors as it will delay their potential outcome payments for outcomes that are realised earlier. The result is that these SIBs also evaluate outcomes in their monitoring process and will have less of a before-after comparison (there is also no baseline measurement for the specific projects), but the achievement of outcomes are monitored on a monthly basis until the end of the period in which outcomes can be achieved. Because it is unclear what the outcomes are compared with, it is simply impossible to prove causality of the intervention, because it is unclear which effects and outcomes can be specifically attributed to the service provider. This makes appropriate evaluation almost impossible. In addition, the burden of proof lies upon the service provider itself, which is, as discussed, problematic in terms of the creation of bias. The good thing compared with the Innovation Fund is that all outcomes in the Fair Chance Fund are objective and measurable, it doesn't contain subjective views of, for instance, teachers.

Ambition East Midlands Through a dedicated 'link worker', young adults of 18-24 will be provided with support to help them find and sustain suitable accommodation and develop skills to enter employment, education, training or volunteering [82]. The project is launched in January 2015 and appears to target all outcomes from the rate card for 340 homeless young people and raised £0.6m for the project [70]. In doing so, it is potentially targeting all three capabilities that are linked to the Fair Chance Fund rate card, Bodily Health, Affiliation and Senses, Thought and Imagination.

St. Basil's The Rewriting Futures programme is developed by St. Basil's to provide a safety net for young people. Through collaborations with involved agencies and having one dedicated individual that is responsible for providing support to service users. This programme is provided for 300 young people and an initial investment of £1m was raised for the project [70]. Support is centred around four pillars. First is to help young people gain accommodation and stability in their lives. Secondly they will be supported to build skills through engaging in volunteering placements, education or training. The third pillar is to sustain full- or part-time work and finally to help the young people build confidence, self-believe and self-esteem [77]. This SIB also will target all different types of outcomes and those outcomes therefore can be compared to the three capabilities that are identified for the full Fair Chance Fund rate card.

Fusion Housing As a housing provider, Fusion Housing combines its expertise with that of employability service providers to provide expertise where needed. The Fusion Housing SIB targets c.300 young people in the area of West Yorkshire and a total of £0.94m was invested upfront [70]. The programme has started specific targets for the 300 young people, being that it aims to support 250 into accommodation, 150 into education and 100 into sustained employment [13]. Through these targets it also shows

that it targets outcomes related to the capabilities of Bodily Health, Affiliation and Senses, Imagination and Thought.

Home Group Also a housing provider, home group uses a supported accommodation model to not only support young people with accommodation, but also support them on-site with additional services. For this SIB they will aim to do that for 230 young people in North East England and they raised an initial investment of £0.5m for that [70]. It is however nowhere stated what this extra support entails and if the SIB also targets other outcomes than the accommodation ones. Therefore it will be assumed that the Home Group SIB will mainly target accommodation outcomes, which can be linked to the Bodily Health capability.

Depaul UK In a programme called ‘Fair Chance, Your Chance’, Depaul UK aims to provide long-term support for young homeless people in Manchester, Rochdale, Oldham and the London Borough of Greenwich. In total it has set out to support 178 young people and an initial investment of £0.63m was raised [70]. With that it is one of the smaller SIBs in the Fair Chance Fund, but it also targets all different outcomes through a long-term and individual approach by support workers. In doing so, the SIB has identified to target all homelessness, education and employment outcomes in the Fair Chance Fund and can therefore also be linked to all three capabilities from the fund.

P3 and CCP This SIB is one of the few where there are two main service providers, P3 and CCP. It combines the expertise of two different service providers, where CCP has its main focus on homelessness, P3 focuses on unlocking the potential in vulnerable young people. Based on those combined services, C3 and CCP intend to transfer 150 young people in Gloucestershire out of homelessness and into employment, education or training [70]. The project aims to realise all three types of outcomes as determined by the Fair Chance Fund rate card [51]. Therefore, this SIB also targets the outcomes that are related to the Bodily Health, Affiliation and Senses, Imagination and Thought capabilities.

Local Solutions The smallest SIB in the Fair Chance Fund is that of Local Solutions, in which £0.55m is raised to support 130 homeless, NEET young people [70]. This SIB also works with two main service providers, Local Solutions and the Oakmere Community College. Together they have developed the specific programme called AIMS (accommodation, intense monitoring and skills) and as with all SIBs in the Fair Chance Fund, it is clear that they aim to support the young people towards accommodation, there is nothing explicit around education and employment. Therefore it will be assumed that the Local Solutions SIB will mainly target accommodation outcomes, which can be linked to the Bodily Health capability.

MHSA

The MHSA SIB is significantly larger than all the Fair Chance Fund SIBs in terms of investment raised (\$3.5m) and aims to provide stable housing for 800 homeless individuals in Massachusetts, US. It is also a rather long project with a duration of six years. The SIB also has an interesting mix of investors where \$1m is invested by philanthropic institutions, while the remaining \$2.5m is invested by private capital such as Santander Bank [32]. The potential outcome payments could lead to a maximum possible return to investors of 5.33% IRR. The SIB has also published its full contract online for transparency, but also show the exact outcome targets. The target of the programme is to achieve sustained housing and is paid \$8.22 per day in which that is successful. This can be up to a maximum total of \$6m or just under 1,000 days of sustained accommodation on average per participant, assuming 800 participants [81]. With this, the aim is to provide adequate shelter for people who didn't have that previously. Having adequate shelter is identified by Nussbaum as being part of the central capability of Bodily Health [47].

The burden of proof for the outcomes lies upon the service provider, but the evaluation is validated by Root Cause, which is a third party to this SIB [33]. The evaluation makes use of matched comparison approach through the availability of validated administrative data. As there is only one outcome that is measured objectively, there are less issues around biases. More important in this case is that the evaluation also looks at the causality of the intervention. The intervention targets 800 homeless individuals, which makes using a comparison group relatively hard, but as governments collect data on homelessness and the status of homeless individuals, using a matched comparison approach by using validated data on the statistics of homeless individuals finding housing, can show the difference with

individuals that have participated in the intervention. It also allows for a full before-after comparison, for both the actual data and the baseline data, but it is unclear whether this actually happens. Therefore, it is unclear to what extent exogenous risks are taken into account in the evaluation of the SIB.

St. Mungo's Broadway and Thames Reach

This SIB is basically a two-part SIB together where St. Mungo's Broadway and Thames Reach both initiated a SIB. In London, a cohort of rough sleepers was identified based on a local database that recognises the rough sleepers in London. A rough sleeper is estimated to cost the UK government around £37K per 5 years, and for the full group of 831 young people in this SIB, the total costs add up to over £30m. For both SIBs, £1.2m in investment was raised. This group is split between the two SIBs and outcomes are measured on five items [70]:

- Reduction in the number of people sleeping out each quarter, compared to a baseline which falls over the years to look for sustained improvement.
- Entry into accommodation and sustainment of tenancy.
- Confirmed returns of foreign nationals without the right to reside in the UK to their home countries.
- Increased employment
- Reduction in the use of Accident and Emergency services.

These outcomes are very diverse and when compared to the Capabilities Approach, they also can be linked to different central capabilities. As established earlier, finding accommodation can be linked to the Bodily Health capability, while employment can be linked to the Affiliation capability. The fifth outcome of a reduction in the use of Accident and Emergency services, can most likely also be linked to the Bodily Health capability as Nussbaum described that capability also as being able to have good health [47]. The remaining outcome of returning foreign nationals without the right to reside in the UK to their home countries is somewhat more difficult however. There is a clear link with the capability of Bodily Integrity as Nussbaum defined that as being able to move freely from place to place [47]. Even though most operationalisation examples have focused on other items for the Bodily Integrity capability, such as the acceptance of homosexuality and safe birth control, this outcome is clearly part of the Bodily Integrity capability. If strictly looked at the Capabilities Approach however, forcing the foreign nationals to return to their home countries actually negatively influences the capabilities of the service users, as this goes against the free will of that person. Therefore, these two SIBs and their outcomes become interesting as this third outcome is defined as positive from the perspective of the stakeholders of the SIB, while it can be seen as negative from the perspective of the Capabilities Approach.

The different outcomes are also evaluated separately. To start with the first three outcomes, they are measured per person and the valuation is based on historical data. This leads to problems with causality, just as it does for the projects in the Fair Chance Fund as it is unclear how much of the outcomes can be attributed to the intervention. Also, the burden of proof for these outcomes are upon the service provider, while the outcomes will be validated by the commissioner. This is less of an issue however, as the objective outcomes create little risk for bias. The final two indicators are evaluated using baseline data and third party evaluators (two different evaluators, one for each outcome). This allows for better evaluation and a full before-after comparison can be done on those outcomes both for the actual data and the baseline data. A complex situation is created though by using different evaluation methods for the different outcomes and involving multiple independent evaluators. Although the criteria of the evaluation design do not speak on the number of different evaluators, it might be more sensible to use only one to avoid unnecessary complexity. The main advantage of using multiple evaluators is that evaluators can be used based on their specific expertise of a subject.

Conclusions Housing/Homelessness

The majority of the SIBs in the Housing and Homelessness section are commissioned by the UK government through the Fair Chance Fund. As these SIBs all only started early 2015, there is also little known on the outcomes or relations within the SIBs. It appears to be the case that all SIBs target the diverse

set of outcomes that is defined by the Fair Chance Fund rate card, although for two SIBs it couldn't be identified to what extent they will actually strive for educational or employment outcomes. These outcomes can be grouped in accommodation, employment and education and can be identified to be related to the Bodily Health, Affiliation and Senses, Imagination and Thought capabilities respectively.

The remaining three SIBs in this category also target accommodation outcomes in providing services for homeless (young) people. For the St. Mungo's Broadway and Thames Reach SIBs there is a more diverse set of outcomes as well, which can also be seen to be in line with the Affiliation and Bodily Integrity outcomes. Interestingly, the outcomes related to the Bodily Integrity capability appear to actually have a negative effect on one's capabilities.

The evaluation methods of the Fair Chance Fund show similar issues to that of the Innovation Fund or Youth Engagement Fund, which are discussed earlier. Nonetheless, the indicators used in the Fair Chance Fund and all Housing/Homelessness SIBs are all objective and therefore can be measured without bias. The other two SIBs both use independent evaluators to validate the evaluation, which is good to have an independent look on the outcomes as well as the processes. The Thames Reach SIB adds unnecessary complexity though by using different evaluation methods and evaluators within one SIB. Overall, the aspect of measuring well-being and looking at the processes of the SIB appear to be missing in all SIBs, even though they are very important aspects of the evaluation.

6.2.5. Child and Family Welfare

In the Child and Family Welfare category, some very big SIBs have been launched as in a total of 8 SIBs, over \$33 million is invested. All of these SIBs are individually launched without the support of a large outcomes fund such as the Fair Chance Fund or Innovation Fund. 5 out of the 8 SIBs have been launched in the UK, while two are launched in the Australia and the final one in Canada. The topic of child and family welfare is a broad topic which shows in the targeted outcomes of the individual SIBs.

Frontline

It is identified that children in foster care settings experience long-term negative effects due to short-term placements and having to deal with separations from their parents on multiple occasions [70]. Therefore in this SIB, 135 homeless families are supported to reduce the length of stay in out-of-home foster care placements for the children of those families [29]. The SIB aims to transfer those families into housing to provide them with a stable home environment and prevent the families from living on the street, but also to reduce the time children spend in out-of-home foster care. The latter is the main goal of the project and the SIB has provided an insight in the way outcome payments are handled. The Cuyahoga County in Ohio, that issued the SIB, will pay \$75 per reduced foster care day and targets to reduce these days with 25% for the 135 families. In a payment scheme, the savings are outlined for the county for different levels of success and while all gross savings will be paid out at lower success levels, the scheme is estimated to realize some net savings for the County as well, starting at a success rate of 25%, but this could run up to \$3.5m in savings at a 50% reduction in out-of-home days [29].

Supporting the families into stable accommodation is a mean to get to the actual targeted outcome, but could still result in increasing the capabilities of the members of those families, as having adequate shelter is part of the central capability of Bodily Health [47]. Children that have to go into foster-care (temporarily), might have a disadvantage on the Affiliation capability, which is also described by Nussbaum as 'Being able to live with and toward others' [47]. Even though, Affiliation is mainly linked to employment [56, 61], it can also be seen as having friends or family around you [6].

With the support of the independent evaluator Case Western Reserve University, the SIB is set up as a randomised control trial [33]. As discussed previously, this will be the most thorough setup towards evaluating causality of a project. As with the SIB in Portugal, the Frontline SIB works with relatively little participants which makes it easier to do a full randomised control trial than with large groups. Especially in cases with family situations like this, the exogenous effects on the outcomes can make huge differences and it remains very difficult to fully prove causality even when a randomised control trial is used. In this SIB, but also in comparable SIBs, using the participant as the unit of analysis for part of the evaluation would become more complex as the family will consist of multiple people and therefore the questionnaire should be used on all members of the family to test how the whole family has perceived the intervention and its outcomes. The most important member of the family will be the child (or children) that are placed out-of-home, but is important to evaluate how the surface is perceived by other members of the family, this is especially important to learn more on the potential unintended

outcomes that the intervention might have realised. This will increase the amount of resources that are needed to do that part of the evaluation though.

Care Asset group

In Birmingham, there are approximately 180 looked after children who are in secure residential care due to repeated breakdown of their foster care situation. The SIB Care Asset Group aims to support 115 young people aged 11 to 15 in residential care to support to foster placements and provide the needed support to make this placement last with wrap-around services [70]. In 4 years, the young people are supported and outcomes are measured firstly by placement in the programme and finding a suitable accommodation, but secondly by sustaining that placement for 52 weeks. Even though the placements are with foster parents and not with necessarily directly with family, this can still be identified as related to the Affiliation capability as is the case in the Frontline SIB [6, 47].

This SIB needs more extensive evaluation, primarily to identify its effects outside of the two indicators that are used for outcome payments. The indicators can be more identified as outputs, whether they participated in the programme, than outcomes. This is interesting, as one of the advantages of SIBs is that evaluation shifts from only inputs/outputs to outcomes and achieved results. Therefore it can be argued that the used indicators are not constructed properly and therefore probably won't follow the process of constructing indicators from a causal model and separating outcomes and outputs. This will make the proposed evaluation method not applicable for this SIB as it will be too extensive. Nonetheless, evaluation of the impact on the participants and evaluating the processes of the SIB remain important and should be added.

Action for Children - Manchester

The target group and social issue for the Action for Children SIB in Manchester, is similar to that of the Care Asset Group, providing Multi-Dimensional Treatment Foster Care where the foster care situation is supported 24 hour per day in a 28-week programme. In total, £1.2 million in capital is raised to support 96 young people who are currently in care. Outcomes are paid based on two main metrics, the length of the placement and the child's well-being [70], however it remains unclear how especially the well-being aspect is being measured in this SIB. As with other SIBs in this category, finding a good and sustained foster care placement can be related to the Affiliation capability, but well-being in general is a core concept of the Capabilities Approach in itself. Unfortunately, there are no insights available in how that well-being is being measured and to what extent this uses the Capabilities Approach or concepts of it.

Historical data was collected from 2007-2008 to use as a baseline for the outcomes. While the collection of the baseline data is in line with the criteria for evaluation as developed in this thesis, the intervention started in 2014 and therefore a lot might have changed since the situation of 2008. Therefore, it should be questioned whether the baseline data was still representative when the intervention was active. Ideally, and the data for the length of the placement would allow that, another run of the baseline data would have been done shorter before the start of the intervention and/or will be done at the end of the intervention. Measuring well-being in this SIB is very much in line with the initial design of the evaluation tool, although the well-being is measured qualitatively. It is not clear how that evaluation will compare with the proposed qualitative design or with the Capabilities Approach.

Sweet Dreams

The only SIB launched in Canada thus far, is aimed to support single mothers and their children, by providing a communal living environment to prevent the children being taken into care. The SIB is very focused and works with a total of 22 mothers with children of 11-15 years old, who are at risk of entering care [70]. Outcomes are measured by one metric, which is that a success can be claimed if a family leaves the communal accommodation and stays together for more than 6 months without any of the children entering care [75]. Although accommodation is provided, results are not measured based on sustained accommodation after the intervention, but on the sustainment of family relations, keeping the family together. As discussed earlier in this section, this can be related to the Affiliation capability.

Even though the SIB is very focused, it uses baseline data for evaluation rather than comparisons groups, but still appropriate. The outcomes are measured by an independent third party and it is stated that by using the baseline data, the added value will be evaluated [33]. This would mean that the outcomes are compared with the baseline, which seems to be an appropriate method. There is

no mention of evaluating the impact on the participants and as with the Frontline SIB, the unit of analysis should be both the mothers and their children for that. Especially with the small number of participants that should not be an issue in terms of resources. Also, there is no mentioning of monitoring or evaluating the processes.

It's All About Me

It has been identified earlier in this paper that SIBs might be useful for service users if they are able to provide long-term funding for service provider. The *It's All About Me* SIB is an excellent example of that, providing funding for the duration of ten years. In total 650 will be aimed to be placed in foster care from their current care settings. The programme will operation nationwide in the UK and uses four outcome metrics. Outcomes will be paid for: the registration of a child, successful placement in foster care of a child, one year spent in placement, and two years spent in placement [70]. Like other SIBs in this category, this SIB aims to give young people a permanent adoptive family and therefore giving them the chance to have family around them, which can be seen as part of the Affiliation capability [6, 47].

Again, the SIB uses an indicator that can be seen as an output rather than an outcome, the registration of a child. The other three indicators are more outcomes and determine success of the programme. They are evaluated using baseline data, while other outcomes are evaluated through a longitudinal study [33]. This is a good addition as unintended outcomes might be picked up through that study. There is something interesting about the duration of a placement as it creates incentive to extend a placement to just past the cut-off time, in order to trigger an outcome payment, while the placement wasn't completely appropriate any more. This should be considered through the evaluation of the processes, and more importantly the evaluation of the well-being of the participants. In this case, both the children as the foster parents should be analysed by the evaluation in order to get full insights into the outcomes of the intervention.

UnitingCare Burnside

In New South Wales, Australia, the service provider UnitingCare Burnside aims to work with 700 children who are at risk of entering out-of-home care [70]. The SIB was launched in August 2013 and in May 2015, the results of the first cohort of 157 children were published and changes have been suggested on the programme for the remainder of the SIB. Primary outcomes for families where at least one child under 6 who has been placed in out-of-home care, are based upon 'restoration' to their family. For families with children at-risk of going into out-of-home care, outcomes are realised if the children remain with the family for at least 12 consecutive months. Out of the 157 children in the first category, 68 were restored to their family. Out of the second category 22 out of the 29 children have remained with their family [76]. As with more SIBs in this category, the outcomes aim to restore or retain a good family relationship, which is in line with the Affiliation capability [6, 47].

The SIB uses a combination of two methods to evaluate the restoration data. For the first three years, a baseline of 25% is used, based on historical data. After the first three years, the restoration rate is determined based on a control group [33]. The use of two different methods is not in line with the designed evaluation tool, but is also confusing as the baseline might suddenly change after three years, while there isn't necessarily something changing in the intervention. An argument to use the control group only after three years is that a control group is still being formed and no reliable results will be found when the group just started, but then the same would go for the participants in the actual intervention. Two solutions would have been better, either doing another test of the baseline data and control for the changes in the baseline to account for exogenous effects, or use the control group for the full duration of the intervention.

Additionally, the SIB will be evaluated qualitatively and on the impact it has on the parents. This will be on top of the regular evaluation and doesn't account for the outcome payments. Even though it is unclear how that evaluation looks like exactly, it is in line with the intentions of the designed evaluation tool in this thesis, but should be defined more clearly.

Benevolent Society

Another, and the largest, SIB in Australia. This time ten million Australian Dollars was raised to support 400 families with children at risk of entering care. The programme targets families with resident children up to 5 years old, who are identified as being at risk of significant harm [70]. The

service's successes are measured on three criteria, the number of entries into care per child, the results of Safety and Risk Assessments and the number of reports per family to the Government's child protection helpline. These are mainly related to living and growing up in a safe environment, which can be seen mostly in Nussbaum's definition of the capability of Bodily Integrity, for instance through exposure to (domestic) violence [6, 12] and living in a non-hazardous physical environment [61].

The outcomes are calculated into a weighted average and compared with a control group for a quasi-experimental setup of the SIB. It is unclear however, how that evaluation takes place and who the evaluator is. It is clear that the service provider is supported to get a better understanding of what success is [33], what could imply that the burden of proof lies upon the service provider. This creates a risk for bias, especially as the outcomes are not all directly objective. Additionally, more should be known on how the control group is formed and evaluated. Other than that, there is little known on the evaluation of this SIB.

Action for Children - Essex

The first SIB that was launched with Action for Children in September 2012, in Essex, UK. Through intensive Multi-Systemic Therapy (MST), which is recognised as an evidence-based programme to improve parenting and rebuild positive family relationships [70]. Over five years, Action for Children works with 380 families to support them to manage families with the primary goal to reduce the number of care days for children in those families. But the SIB also measures five additional outcomes of the programme [72]:

- Youth offending records pre, during and post MST
- Educational attendance and where possible attainment
- Family outcomes, indicating 'distance traveled' by the family
- Satisfaction with the service as measured by qualitative family feedback post intervention
- Psychological wellbeing, measured through the Strengths and Difficulties Questionnaire by the young person and their carer pre- and post-intervention

The outcomes concerning health can be seen as in line with the Bodily Health capability and educational attainment is seen as being in line with the capability Senses, Imagination and Thought. Broader family related outcomes are harder to compare with individual capabilities, but based on improved family relationships, these could be argued to be in line with that of the Affiliation capability.

Evaluating the outcomes of the Essex SIB is more complex as not all outcomes are objective or based on hard data. The softer outcomes are more in line with other aspects of the designed evaluation tool in this thesis though. Family outcomes and psychological well-being are somewhat in line with the proposed evaluation of the participant's well-being and are as proposed, performed both pre- and post-intervention. Additionally, a qualitative evaluation concerning feedback on the process and the satisfaction is used, which is in line with the evaluation of the processes of the SIB. The hard outcomes are evaluated using historical baseline data, rather than current baseline data. Most important though, the evaluation is still performed by the stakeholders of the SIB. Even though the setup appears to be broad, particularly the qualitative and well-being evaluation should be performed by third party evaluators in order to create the appropriate validity for the evaluation.

Conclusions Child and Family Welfare

The largest group of individually launched SIBs falls into this Child and Family Welfare category. Most SIBs are aiming to keep children with their families or to give them a stable home environment with foster parents. These outcomes are linked to the Affiliation capability mainly, although to a lesser extent some SIBs can be seen to focus on the Bodily Health capability as well. Finally, the outcomes of the Essex SIB can also be seen to be in line with the capability of Senses, Imagination and Thought. Making it the third capability that outcomes are linked to within this category.

The Essex SIB is a good example of a more complete evaluation setup, even though all parts will be used for outcome payments. However, the SIB is not using a third party as the evaluator, which creates the risk of bias in the outcomes. The other SIBs use either baseline data or sometimes control groups for evaluation. Specifically for complex SIBs such as these, control groups can be useful, although most

outcomes that are defined appear to be relatively objective and therefore could be measured through baseline data as well. In other categories of SIBs, the participants are usually individuals, however in this category, clearly the whole family participates (or in some cases the foster parents). This should be included in the evaluation as they might have different perspectives on the outcomes and the processes. Therefore, all participants should be evaluated individually, even if the participants are multiple people in this case.

6.2.6. Criminal Justice

SIBs targeting service users with a criminal history were the first to be launched (the first ever SIB in Peterborough falls in this category) and appear to have been on a much larger scale per SIB than other categories, although there haven't been many different SIB, with close to 16K of targeted service users in only 4 SIBs. This is also the only category with multiple SIBs where the majority has not been launched in the UK, with 3 out of 4 being launched in the US, of which two in New York.

Roca

Probably the largest SIB to date is that in Massachusetts where service provider Roca aims to support 929 at-risk young men of 17 to 24 years old. A total of \$21m in investment was raised for this seven year programme [31]. The service will target young men who are or have been in probation or custody and aims to reduce recidivism and increase employment through an intensive support service. Outcomes are first of all paid by a reduction in the days of incarceration, which comprises most of the potential outcome payments and can run up to \$27m if the days of incarceration are decreased by 70%. Other payments are made for increasing the engagement of the service users with the service and for an increase in employment versus a control group [31]. From a capabilities perspective, an increase in employment is clearly linked to the Affiliation capability. Reduction in reoffending is harder to directly relate to a single capability. There is obviously a relation with the Bodily Integrity capability as this includes *being secure against violent assault, including sexual assault and domestic violence* [47], but it could be argued that mainly other people in that environment are beneficiaries of the reduction in offending. However, if strictly looking the Bodily Integrity capability also includes *Being able to move freely from place to place* [47], which is obviously reduced when someone is in prison. Another potentially related capability is that of Practical Reason (being able to form a conception of the good), but proxies for that have mainly been used based on mental health/capacity [61]. It appears that in its essence, Bodily Integrity would be a more relevant capability for reduced incarceration.

The mix of outcomes is interesting as they are very different in their essence and are also evaluated differently. While a decrease in incarceration and an increase in employment are clear and objective outcomes, the increase in job readiness (based on engagement with the service provider) is more of an output. Interestingly, they treat the employment and incarceration outcomes differently in terms of evaluation. Where incarceration is evaluated via a randomised control trial, the employment figures are compared to a administrative baseline. Both could be easily measured by baseline data, seeing which statistics need to be collected, but when a randomised control trial is possible for incarceration, it would have made sense to use the same groups to measure employment statistics. Finally the engagement with the service provider has a high risk of bias as it doesn't state what the quality of the engagement needs to be and therefore can be easily seen as an output that can be increased more easily by the service provider in order to trigger outcome payments. All metrics are evaluated by a third party evaluator [33], which increases validity, but doesn't reduce the risk of bias in the specific metric as this can be skewed directly by the service provider. In that case especially, extra evaluation is needed for the well-being of participants and to qualitatively test the processes of the SIB in order to evaluate what the level of engagement has been. There is no mention of that however.

Center for Employment Opportunities

In 2013, this SIB was announced by the New York State to support 2,000 formerly incarcerated individuals with the aim of a reduction in days of incarceration, increased employment and improving public safety. The Center for Employment Opportunities provides employment services, including training and job placement. The programme will run for 5.5 years and \$13.5m in investment was raised, for the first time through the network of a commercial wealth management platform to its investors [74]. The three targeted outcomes are an increase employment in the fourth quarter following the release from prison, a reduction in the number of days incarcerated following release from prison and to maximize

participant engagement in transitional jobs [45]. The first and the third are clearly related to the Affiliation capability through the employment outcomes. The reduced recidivism is argued to be related to the Bodily Integrity capability in the discussion of the Roca SIB.

The SIB is set up as a randomised control trial. When done well, this is the optimal method for validity of the outcomes and to prove causality of the intervention. The evaluation is however performed by the commissioner, and even though it will be validated by an independent organisation [33], the commissioner is also most likely to assign the participants randomly. This might create a risk of bias, although that doesn't necessarily have to be the case when the randomised control trial is designed well. The metric around the transitional job is not evaluated through the randomised control trial as that won't deliver results for the control group, the transitional jobs are a specific outcome of the intervention. These are compared with a baseline created before the intervention was started based on administrative data. There is no mentioning of any evaluation of the well-being of participants, or the processes of the SIB.

Rikers Island

In New York City, 50% of youth that enters jail, will return within a year. Therefore, the SIB in the New York City Rikers Island prison, aims to reduce the reoffending rate for that group [31]. All 16-18 year olds that enter the Rikers Island prison and stay more than 4 days (c.3000 per year), will be targeted and their reoffending rate is compared to that of a control group. The only outcome target that is being targeted is a reduction in reoffending days, which is defined as being related to the Bodily Integrity capability. Nearly two years after its launch, this SIB has been discontinued, due to a lack of success [8], resulting in a full loss for the investors. What the exact results were in terms of reoffending by the service users is unclear.

The evaluation was performed by a third party organisation, but the exact methodology is unknown. It is clear however that there was a monitoring aspect to the evaluation as the programme was discontinued based on a lack of success, which means that success was measured as the intervention was in place. This is in line with the design of the evaluation tool, to not only monitor inputs and outputs, but also monitor the development of the outcomes and have that part of the negotiations for future implementation. In this specific case the negotiations have led to the decision to cease the intervention. Although that might be a shame, this also means that most likely the monitoring tool has worked and was used for the right purpose.

The One* Service

The SIB in Peterborough was the first SIB to be launched, in September 2010. In total, £5m was raised to provide services to 3,000 short-sentence offender. All service users that have a short-term sentence and are released from the Peterborough prison will receive wrap-around services to reduce reoffending, which will be measured against a control group. Investors will receive a return if the reoffending has been reduced by more than 7.5% compared to the control group [70]. Although the SIB appeared to be on course to make that target of 7.5% in 2014, the SIB will be discontinued in 2015 due to a change in policy focus by the UK government [73]. The targeted outcomes of a reduction in reoffending can be linked to the capability Bodily Integrity.

As the intervention used a control group, which was in a different geographical area, the SIB was set up as a quasi-experimental project. Outcomes were compared with the control group which decreases the exogenous risks to the validity of the evaluation. There was a clear monitoring function both on the processes and the inputs/outputs. The processes needed to be monitored as the intervention wasn't a fixed design and could be flexible based on the participant's needs. Therefore the need for qualitative evaluation of the processes is even higher in order to evaluate which approaches have worked and which have not. Also, as the groups were split in different cohorts, evaluation took place after every cohort as was suggested in the evaluation design of this thesis. This has led to several progress reports. The project was discontinued early, but that wasn't the result of a correct monitoring approach but of a separate policy decision made by central government, which is technically also the commissioner.

Conclusions Criminal Justice

Four SIBs so far, although unfortunately two of those have been discontinued early. The SIB in Peterborough was discontinued for other reasons than its outcomes, but the Rikers Island SIB had to be discontinued because of a lack of success. Even though the originators argue that the provided learnings

still make that SIB a success [8], also because the SIB has been able to provide services to the target population and test the type of services at no costs to the taxpayer as all costs are covered by private investors that have taken up the risk of having no outcome payments.

In general, all SIBs have been targeting the Bodily Integrity capability. It could be argued that also the Practical Reason is relevant in some cases, but that is more related to mental health/capacity than incarceration. In half or the SIB, also employment outcomes are target, resulting in an attachment to the Affiliation capability as well. For SIBs like this there is also an argument to be made that reducing criminal activity for the target population also has effects on their surroundings and thus on the citizens living in the proximity of those SIBs. This could therefore also be relevant for the Bodily Integrity capabilities of people outside of the main target population. For measurement purposes, it will be hard however to assess the improvement of the capabilities of each citizen that lives in the areas that are impacted. Therefore assessment of the capabilities is probably best to focus on the service users themselves.

Two of the four SIBs are partially set up as a randomised control trial, but in both cases the randomised control trial is only used for part of the outcome indicators, while others are evaluated using baseline data. This is an interesting mix, but it would have made more sense to use either one of those methodologies, rather than a mix of them. The One* Service SIB was set up as a quasi-experiment, while the evaluation methodology for the Rikers Island SIB is unknown. It is clear though that the two latter SIBs have been using a monitoring aspect to evaluate both inputs/outputs and the outcomes, which is in line with the design in this thesis and shows that that does not have to be an issue in terms of resources, as most of the SIBs in this category are relatively large in the number of participants. It appears though that the well-being of the participants is not taken into account, neither are the unintended outcomes looked at. In Peterborough, it became clear that unintended outcomes can play a significant role in especially this category [57] as the SIBs will require the participants to go through a lifestyle change, which also might effect their surroundings. Therefore, the evaluation aspect of well-being and the qualitative evaluation should play a bigger role in these SIBs than they currently appear to do.

6.2.7. Health

So far it has been proven hard to launch a SIB focused on health, which is shown by the fact that only one SIB in this category has been launched and also in 2015. Although health can be seen as one of the growing concerns for governments over the world, it is also a sector where it is much harder to find an evidence-base for services or find a single government that is able to pay for outcomes, as most outcomes will have inter-departmental effects.

Social prescribing

The first ‘solution’ to launching a SIB that targets health outcomes was found in Newcastle, UK. 70% of the costs of the NHS is spent on people that suffer from long-term health conditions such as lung disease, diabetes and asthma [70], which are therefore a major issue for the NHS. For this target group, services based on social prescribing are offered to realise more sustained healthy behaviour and improved self-care. Social prescribing works as a non-medical intervention ranging from interventions such as physical activity to social interaction and welfare rights advice [97]. Outcomes are target based on an improvement of well-being (measured every six months using a ‘Wellbeing Star’) and the reduction in secondary care costs in comparison to a matched cohort [70]. Although an improvement in well-being can be related to the more general Capabilities Approach, the targeted improvements in health and reduction of secondary care costs can clearly be linked to the capability of Bodily Health which can be measured by someone’s physical shape [92] or contacts with doctors and nurses [12].

The health of people, and particularly issues that are treated through social interventions, can be affected significantly by exogenous effects. This should therefore be included in the evaluation, however as this appears to be a very specific target group, it doesn’t appear to be suited to collect baseline data for the outcome metrics, but it will also be troublesome to create a quasi-experimental project as it will be hard to find a control group that has the same characteristics. The only valid way to evaluate this SIB appears to be by using a randomised control trial. Although that might be troublesome ethically, from an evaluation point of view, it does appear to be most sensible. In this case a matched cohort was created, but it is unclear how that was done in detail, it might be hard to actually compare those groups. Finally, the outcomes are also based on an improved well-being, where a specific approach is used to

measure the well-being of participants. This is an alternative for the Capabilities Approach, although using the Capabilities Approach as a whole, would also work to evaluate the SIB and the evaluation tool design would still hold for this SIB. It is unclear how and how often the project is monitored and whether a further qualitative approach was included in the evaluation.

Conclusions Health

As launching SIBs in the health category has proven to be difficult, there is just one SIB. As might be expected, the outcomes are linked to the Bodily Health capability, which can also be expected of potentially future SIBs in health.

6.2.8. Adults with Complex Needs

People with complex needs are not only harder to help than their peers with less complex or no needs, it is also harder to find commissioners that are willing to pay for outcomes, as those complex needs and therefore their costs can most likely be related to a variety of governmental departments. As with health, it has therefore been hard to launch a SIB for this particular group, but also in this category, the first SIB was launched in 2015 in Worcestershire, UK.

Loneliness

Older people are more vulnerable to loneliness and social isolation, which can lead to increase health and social care costs due to increased risks to depression, dementia and strokes or heart disease. The Loneliness SIB in Worcestershire aims to work with 3,000 adults of 50 years and over that have are lonely based on a loneliness measurement scale. Services aim to integrate the service users back into communities through several activities. The outcome that is targeted is that of a reduction of their loneliness score in 6 and 18 months [70]. Although this outcome might improve the health of the service user, this is not explicitly targeted. Loneliness in itself however, can be linked to the Affiliation capability, by having the ability to engage in social interaction [47] or having friends and family around [6]. Loneliness, and especially not being able to engage in social activities, can also be related to the Play capability through the ability to *Being able to laugh, to play, to enjoy recreational activities* [47].

By targeting the loneliness score without a baseline or control group, the underlying assumption is that the loneliness score would not improve without the intervention, which is hard to prove completely, but not unreasonable to assume in this specific case. Measuring the loneliness in itself is already part of a broader evaluation of the well-being of the participants and this aspect will therefore include both the outcome indicators and well-being evaluation. As there is a single outcome and no comparison, evaluating and monitoring the processes will be extremely important and should be part of careful consideration by the stakeholders on a frequent basis in order to really overthink the approach taken and whether that is successful. The loneliness measurement should definitely be performed by an independent evaluator as it is highly subjective and prone to socially desirable answers, therefore the evaluation should be used carefully and not by any of the involved stakeholders, to avoid biased results. In general this approach has a significant risk of bias towards outcome payments and it should be considered to also use other more objective outcomes. An example could be evaluating the number of engagements in social activities separately.

Conclusions Adults with Complex Needs

Although there is only one SIB in this category, the SIB appears to target two capabilities with the main one being the Affiliation capability, but also the Play capability plays a role. The Affiliation capability is more in line with that of the Housing and homelessness SIBs than for instance that of the workforce development SIBs that are more focused on the employment aspect of Affiliation.

6.2.9. Conclusions on SIB case study

While some categories have seen more SIBs than others, the current SIBs do give a broader sense of the type of outcomes that SIBs might be targeting and to what extent those outcomes can be related to central capabilities of the Capabilities Approach. The identified capabilities might not be directly able to replace outcome measurement that is taking place within the SIBs, but getting a broader sense of the capabilities helps in getting a sense of where potential improvements in one's personal capabilities might be expected due to the interventions. By combining the related capabilities and setting up a measurement framework, SIBs potentially can be measured not only on their internal success, but they

can also be compared to other SIBs to see which SIBs are really improving the capabilities and the well-being of the service users.

Analysing these SIBs will inform all aspects of the tool from various perspectives, not just the focus on the Capabilities Approach. Analysing the SIBs that are or have been active, provide insights into the SIB systems that have been used and the evaluation method that has been used so far. The systems will lead to three typologies of the SIB system, which is described later in these conclusions. Looking at the current SIBs it shows that they are diverse in every aspect. Not just the social issues that they are solving are different, or the methods that are being used to do so, but also the setup of the SIB has been varied. This also leads to a wide range of evaluation methods. Even though, different doesn't necessarily mean bad, it appears to be that not all cases are evaluating to the standards as this thesis defines them. A great example of this is the tariff model that is being used in a total of 23 SIBs. This methodology has very high practical achievability, as it makes it easier to construct a SIB, but the validity of the evaluation also becomes much more troublesome. It should be of concern that half of the SIBs so far have been launched through this method. Even though volume can drive the growth of SIBs significantly, validity and trust in evaluation are crucial for gaining sustained traction and improving the impact that services have. Examples the other way around can be seen too, the discontinuation of especially the Rikers Island SIB is something that is encouraging from an evaluation perspective (although, be assured that it is in no way encouraging that projects that aim to reach social impact do not reach their goals). When a conscious decision is being made based on monitoring of the outcomes as was done in this case, it shows that the KPIs have been sufficiently measurable and have been discussed amongst stakeholders to allow such decisions. Independent of the exact methodology that is being used for the evaluation of a SIB, it is important that it is up to a standard that decisions like that are at least possible to make.

The capabilities that are being identified throughout the different categories will serve as a basis for the framework and questions will be developed for those capabilities initially. A capability that comes through in a lot of SIBs is that of Affiliation. The Affiliation capability comes out through the aim of improving employment outcomes for service users, but also by improving people's home situation through the relationships with family and friends. This means that questions on the Affiliation capability should capture both those aspects. Another capability that comes through a lot, especially in the workforce development and education and early years categories, is that of Senses, Imagination and Thought. In all cases this outcomes are related through educational attainment of the service users, albeit at different stages of service user's life (for instance in primary versus secondary education). The final two capabilities that appear to come through in several SIBs are covering Bodily Health on health related incomes and the Bodily Integrity on outcomes that are related to criminal justice SIBs and reducing criminality and incarceration. In two SIBs, arguments could be made that there is a slightly smaller relation to the Practical Reason and Control over one's Environment, which should therefore be taken into account, although they do not play a dominant role in the SIBs. An overview of the identified capabilities can be seen in table 6.6.

Another important feature is the methodology that will be used to evaluate the SIBs, or in other words, the way the SIB is set up. An overview of the different methodologies can be found in table 6.7. The predominant method appears to be the tariff model, mainly because the three larger SIB funds can be seen as using this method, although the name might need explanation. For this group of SIBs the commissioner has identified the indicators and the valuation of those. Although the valuation is based on historical data and therefore could be argued to be a matched comparison, there is no insight in this and evaluation for the SIBs doesn't take the baseline into account during or after the SIB. Therefore it should be categorised differently from SIBs that actively use baseline data. In total seven SIBs have used a control group in any form.

This overview does show that even though part of the SIBs use control groups, this is not widely spread. Most likely this is due to the complexity of using control groups, while SIBs also need to be practically achievable. The SIBs that have been set up as a randomised control trial are also mainly smaller SIBs in the number of participants, which makes the use of control groups easier. When the per person SIBs are accounted for as using baseline data, as they technically are using baseline data, it is clear that the majority of SIBs have used some form of baseline data. While for four SIBs the evaluation method is known, no SIB can be identified as not being evaluated at all. As the three SIB funds lay the burden of proof upon the service provider (or the intermediary if there is any), a large part of SIBs are seen to having internal evaluations. For the other SIBs however, a large numbers are

Capabilities that are aimed to be improved through existing SIBs

Capability	Number of SIBs	SIBs
Life	1	Action for Children Manchester
Bodily Health	14	Ambition East Midlands, St. Basil's, Fusion Housing, Home Group, Depaul UK, P3 and CCP, Local Solutions, MHSA, St. Mungo's Broadway and Thames Reach, Frontline, Care Asset Group, Action for Children Manchester, Action for Children Essex, Social prescribing
Bodily Integrity	7	St. Mungo's Broadway and Thames Reach, Action for Children Manchester, Benevolent Society, Roca, Center for Employment Opportunities, Rikers Island, The One* Service
Senses, Imagination, Thought	24	Career Connect (2x), BEST Network, Tomorrow's People, Community Links, Adviza, Teens and Toddlers (2x), 3SC Capitalise, Prevista (2x), Juvat, Futureshapers Sheffield, Junior Code Academy, Girl's Education, Chicago Public Schools, Utah School District YMCA & Little Scholars, Ambition East Midlands, St. Basil's, Fusion Housing, Depaul UK, P3 and CCP, Action for Children Manchester, Action for Children Essex
Emotions	1	Action for Children Manchester
Practical Reason	2	Action for Children Manchester, Roca
Affiliation	31	Career Connect (2x), BEST Network, Tomorrow's People, Perth YMCA, Adviza, Teens and Toddlers (2x), 3SC Capitalise, Prevista (2x), Juvat, Buzinezzclub, Duo for a Job, Futureshapers Sheffield, Junior Code Academy, Girl's Education, St. Basil's, Fusion Housing, Depaul UK, P3 and CCP, Frontline, Care Asset Group, Action for Children Manchester, Sweet Dreams, It's All About Me, UnitingCare Burnside, Action for Children Essex, Roca, Center for Employment Opportunities, Loneliness
Other Species	1	Action for Children Manchester
Play	2	Action for Children Manchester, Loneliness
Control over one's Environment	1	Action for Children Manchester

Evaluation methodologies on existing SIBs

Type	Number of SIBs	SIBs
Randomised Control Trial	3	Junior Code Academy, Frontline, Roca (partially), Center for Employment Opportunities
Quasi-Experimental	4	Duo for a Job, UnitingCare Burnside (partially), Benevolent Society, The One* Service
Matched comparison	11	Juvat, Utah School District YMCA & Little Scholars, MHSA, St. Mungo's Broadway and Thames Reach, Action for Children Manchester, Sweet Dreams, It's All About Me, UnitingCare Burnside (partially), Action for Children Essex, Roca (partially), Social Prescribing
Tariff model	23	Career Connect (2x), Nottinghamshire Futures, BEST Network, Tomorrow's People, Community Links, Perth YMCA, Adviza, Teens and Toddlers (2x), 3SC Capitalise, Prevista (2x), Juvat, Futureshapers Sheffield, Ambition East Midlands, St. Basil's, Fusion Housing, Home Group, Depaul UK, P3 and CCP, Local Solutions, Loneliness
Unknown	4	Girl's Education, Chicago Public Schools, Care Asset Group, Rikers Island

using independent evaluators. This is advisable for many reasons that have all been stated previously, but the existing SIBs show that using an independent evaluator is also practically achievable and it should therefore be advised to always do so.

Only few SIBs actually use qualitative evaluation or evaluate the well-being of participants. Interestingly, the SIBs that do so, mainly do so because they have set indicators that require those specific parts of evaluation. Otherwise stakeholders don't have direct incentives to evaluate the perceptions of participants. For a lot of social issues, it can be argued that the defined outcomes are related to broader positive social impact and therefore will also have a positive impact on the individuals. However, it still seems odd that this isn't actually checked with the participants as they are the ones who's live interventions are attempting to improve. No SIB can be seen using the Capabilities Approach and it is unclear what exact methodologies are being used so far. Therefore, the Capabilities Approach can add a crucial aspect to the SIBs to create a standard for evaluating well-being, that is comparable, valid and will show the true effect SIBs have.

Even though there is little actual proof of the use of qualitative evaluation as well as active monitoring. Several examples show that both are being used and imply that especially monitoring is being used more frequently, as shown by the Rikers Island example and some of the Innovation Fund SIBs. This is not surprising as most SIBs deal with several stakeholders who all are interested in positive outcomes, although maybe for different reasons. Therefore especially investors will want to ensure that the intervention actually is on track to achieve those outcomes and adapt the intervention where needed to optimise results. This can be seen by other typical investment funds as well when investing in public companies for instance, regular stockholder's meetings and quarterly financial statements are used to indicate progress. Other items will be reported, but it can be expected that all stakeholders engage in frequent monitoring activities, based upon the examples that can be identified.

SIB typologies

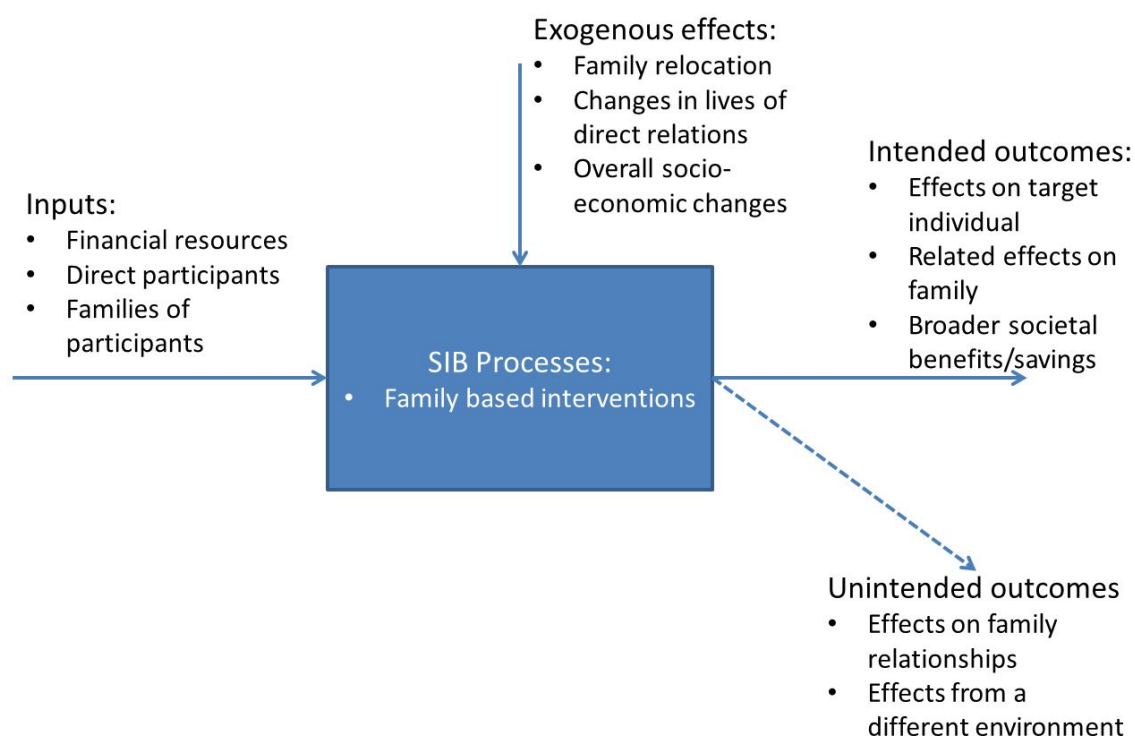
As SIBs are very different, both in setup, in social issue and in stakeholders, to better look at the evaluation tool, three typologies are designed based upon the above case study. These typologies are broader than the earlier categorisation based on social issue and intend to give a better overview on what an evaluation tool needs to meet in order to sufficiently evaluate those types of SIBs.

Although causal diagrams will still be very different within the typologies, based on the exact targets

of the SIB and the intervention that is being used, a more concrete system can be designed with general examples of each part of the system for each of the three types of SIBs.

SIBs where a participant consists of multiple people The first typology is mainly based on the child and family welfare category and will predominantly focus on family issues. In some cases, participants are named as 20 families, but each family will consist of two or more members. Usually one person will be the main target of the intervention (such as the child), but in these cases the rest of the family are directly affected as well. Therefore the effects on them should be accounted for just as well as the main target. This will also mean that the qualitative evaluation and the quantitative well-being evaluation will take up significantly more resources as the number of questionnaires that need to be done will at least double. The qualitative evaluation also should take into account how relationships within the family have been affected by the intervention.

In terms of the system of these SIBs, it will be clear that effects that take place for the rest of the family should be taken into account as intended outcomes and should be included in the indicators, rather than being part of unintended outcomes and being exogenous to the system. However, not all effects on the family can be worked towards and therefore some will still be part of the unintended outcomes, while some effects from the family can still be seen as exogenous. When this is translated into a system, it would look like figure 6.1.

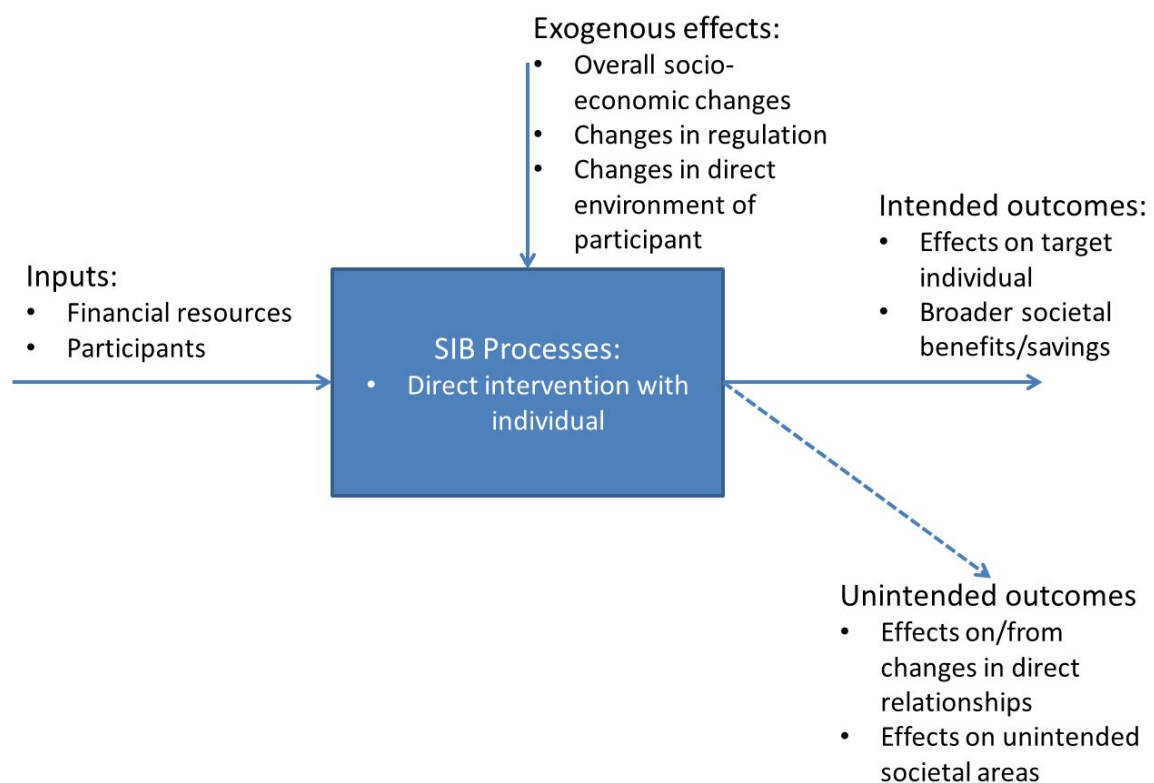


An example system of the first typology

As with regular SIBs, the inputs will be the financial resources and the participants, but in the family environment, also the direct family members will be part of the inputs as they will engage more significantly with the SIB as they would in other SIBs. All SIBs can be influenced by the exogenous effects of broader socio-economic effects, such as a financial crisis. Family related effects can be more severe for these SIBs though. Changes in their lives are not part of the SIB for instance, but can have significant effects. When the aim is to make sure children stay with their families, a parent becoming addicted to drugs (or vice versa, getting off drugs), can have significant effects on the outcomes, the same thing will obviously be for the employment status of parents. These things might not be within reach of the SIB though and should therefore be treated as exogenous. The intended outcomes will be directly related to the main participants, but there will also be intended outcomes on a family level as opposed to other SIBs. Finally, there will obviously be broader outcomes such as savings for the government or the creation of public value. Unintended outcomes for these situations might concern the

family as well, placing a child with a foster family or from an out-of-home care placement back to his original family can potentially disrupt that family, which could have both positive and negative effects. The same will be for the targeted individuals as they might be placed into a different environment altogether, in terms of area, school, group of peers, which can also cause both positive and negative unintended outcomes.

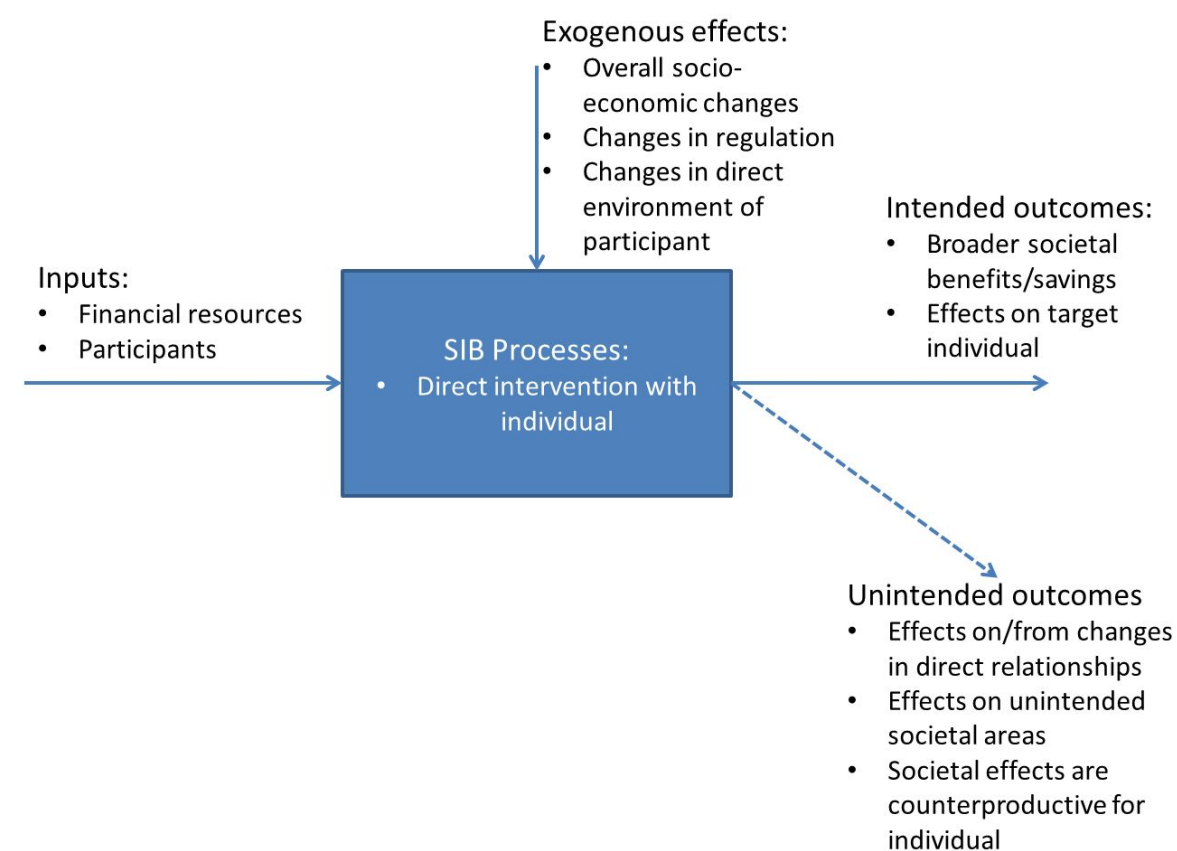
SIBs that target social issues that are primarily individual Social issues are in their essence issues that involve the broader society. However, the way they are targeted can be focused on an individual or on the broader society. When improving employment or educational outcomes, that first and foremost has effect on the individual and only later on society. In these cases, causality between the ultimate goal (i.e. reducing overall unemployment) and the intervention becomes harder as it is more open to exogenous effects as individual success doesn't necessarily mean also broader success is achieved. However, as outcomes are intended to be achieved on an individual level, it will be easier to see the unintended outcomes as they will be achieved on an individual level as well and can be measured through evaluating the well-being on a broader level than just on the aspect that the intervention is trying to improve.



An example system of the second typology

When comparing the system of this SIB as presented in figure 6.2 to the system that targets families as above, the interventions will obviously target single individuals rather than families or groups. This will also mean that the families of participants are not considered inputs in the SIB anymore, as is the case for the intended outcomes. Exogenous effects will still be broader socio-economic changes as they are in every SIB. In this case, changes in regulation will have a stronger effect than in other SIBs as for educational SIBs, a change in the way school exams are taken or in the criteria for achieving a certain grade, will directly affect the outcomes for those individuals. Finally, there can still be direct changes in the environment of a participant, family related or else, that can have an effect on the way a participant engages or is able to achieve outcomes. Unintended outcomes are most likely to occur either from the changes in the relationships of the participant as an effect of the SIB. Also, effects can take place on societal areas that were not foreseen. Having a job rather than being unemployed can lead to effects on capabilities such as Bodily Health and Play, while that is not directly the intention of the SIB.

SIBs that target social issues that are primarily broader societal issues Finally, the third typology of a SIB is one where the outcomes that are targeted are related to broader societal issues, but do not directly lead to a direct solution for the problems of an individual. This is the opposite of the previous typology. An example of such a SIB is where the aim is to reduce recidivism. While reducing recidivism will tackle broader issues around criminality, it doesn't directly solve problems for the individual, which might lie deeper in unemployment, mental or family problems. It can even make things worse, when a drugs dealer is stopped from dealing drugs, a societal problem will be solved, but the individual now has lost his/her source of income and this might cause frictions with the family or makes that he/she can't afford rent anymore. Solving the societal problem therefore might lead to all sorts of unintended outcomes that weren't foreseen and will be different for all interventions. Therefore, the unintended outcomes in these cases are even more important to monitor and evaluate and try to prevent from happening. In setting up such a SIB, it should be considered whether indicators should also be focused on more individual topics. Also, when designing the causal model, it becomes more important to overthink the unintended outcomes and make sure that the intervention tries to at least work to solve the negative unintended outcomes that can potentially arise.



An example system of the third typology

The system of this third typology is set out in figure 6.3. The differences with the system as displayed in figure 6.2 might be subtle, but can make a big difference for the perceptions of the stakeholders and the eventual outcomes of the SIB. Therefore, the inputs and the exogenous effects are similar to the previous SIB typology. The intended outcomes can also be classified somewhat similar, although in this case the broader societal benefits/savings probably outweigh the individual effects. The main difference is that there is an additional unintended outcome, which is that the societal effects won't necessarily have a positive effect on the individual as well, as discussed in the previous paragraph.

6.3. Conclusions SIB Analysis

This section will look at the conclusions from both the interviews and the analysis and look at the consequences for the redesign of the evaluation tool in the next chapter. As the evaluation tool has two main aspects, monitoring and evaluation, it will be discussed what the conclusion will mean for those specific aspects.

First, some overall conclusions. Both monitoring and evaluation appears to take place in SIBs as far as can be seen, the interviewed stakeholders also acknowledge the importance of it. It is not always clear which evaluation methodology is used for SIBs, but in a significant number of SIBs, independent evaluators are used to evaluate the outcomes and the existence of a third party in the SIB is also being seen as important by the interviewed stakeholders. The majority of SIBs that don't use an independent evaluator are using the tariff model, which can be seen as a special model in itself as it requires much less evaluation.

That tariff model is also the most used evaluation model, due to the three funds that have been launched by the UK central government, which all use the same methodologies. Even though it could be argued that those SIBs use historical data to set the tariffs, this is completely done by the government and the baseline is not part of the evaluation, nor is it different for the different SIBs within a fund. Technically though they can be seen as using matched comparison data, which makes that approach the far majority for the existing SIBs. Some have used control groups, or even a randomised control trial. Even though that would be a more valid approach in evaluating exogenous effects and looking at causality, using baseline data is still most used and would be more cost efficient. Also, the interviewed stakeholders do not appear to be convinced that using control groups adds significant value.

From the analysis, three typologies of the SIBs are being made, independent of the social issue they are targeting. As is shown in the previous section, these typologies have subtle differences when looking at the SIB system and will probably look more different when looking at the intervention itself. However, they do not appear to show the need for significantly different evaluation methodologies. The typologies do re-emphasise the need for tracking the unintended outcomes as they can play a role in the actual impact that a SIB eventually has.

Evaluating well-being and using the Capabilities Approach to do so, was seen as an interesting addition by the interviewees, although there were worries whether there were incentives to actually include that aspect in the SIB. In the SIB analysis, the capabilities were analysed more in-depth to see whether the goals of the SIBs are in line with the Capabilities Approach at all and if so, how they compare with the set of central capabilities. In general it can be concluded that the goals of the SIBs can definitely be aligned with certain capabilities. It also becomes clear that the SIBs are targeting certain capabilities more often, such as Affiliation and Senses, Imagination and Thought. While there was one SIB that looked at overall well-being and is categorised on all ten central capabilities, it is not directly affiliated with those capabilities. Leaving four capabilities (Life, Emotions, Other Species and Control over one's environment) somewhat 'unused' in the SIBs. Looking at the content of those capabilities it also seems unlikely that they will be targeted frequently, if at all, in future SIBs. Therefore, when looking at the capabilities that SIBs are trying to improve, six capabilities remain. When evaluating the well-being of service users, it might be more efficient to only look at those six capabilities, especially when looking at the comments on the practical achievability of the evaluation.

Looking at the monitoring aspect of the SIBs, it is clear in some SIBs that monitoring is actively being used for decision making and is also recognised by the different stakeholders as an important aspect of the SIB. Based on this chapter though, predominantly interviews have suggested that monitoring mainly focuses on the data of participants, their numbers, status and the outcomes that have been realised. The processes of the intervention can be discussed in the meetings and be part of the negotiation, but there appears to be no benefit from actively monitoring those processes due to the needed expertise and the expected trust within the group of stakeholders. This is clearly different from the initial suggestion as provided in this thesis. Based on the interviews it can be questioned however to what extent this would provide learnings for the SIB as a process or system, or whether it provides learning purely for the service provider. While the latter remains important, interviews suggest that stakeholders agree on making this the responsibility of the service provider, making it unlikely to be implemented. This combination of arguments therefore lead to the conclusion that this aspect of monitoring can potentially be scrapped from the evaluation tool.

It was also clear that when done correctly, monitoring could lead to adaptive decision making. While the interviewees had concerns around changing standing agreements, one SIB has proven that

monitoring has actually led to changing the decisions. In the case of the Rikers Island SIB, the SIB was discontinued based on a lack of success in the first stages of the programme. This is a good example of monitoring taking place and it has been established that several other systems definitely make use of a monitoring aspect in their evaluation. Albeit a rather radical example of adaptive decision making, the Rikers Island shows that stakeholders within a SIB can put that into practice. If written statements are anything to go by, it appears that all stakeholders have agreed on the early termination of the SIB and are actually satisfied with the learnings that they can take away from the SIB. This re-emphasises the value of monitoring for both current and future SIBs, showing that monitoring in itself should take a central place in the evaluation tool and should be done on a frequent basis.

The evaluation aspect of the tool should focus on evaluating the indicators, the intended outcomes and the revenue that follows the outcomes. As the revenue can follow directly from evaluating the outcomes, that is less part of the design of this evaluation tool as in most cases, the valuation of outcomes is agreed upon beforehand and is seen as a task for the commissioner. The other two parts of evaluation should be adapted based on the information from this chapter though. The evaluation of the impact on participants, using the capabilities approach, can be shortened based on the analysis of the capabilities in this chapter. This will make the questionnaire shorter to process, therefore the evaluation cheaper, and it will make it easier for participants to fill in the questionnaire, improving response rates and therefore improving the validity of the tool. Therefore, the questionnaire should be redesigned based on the six most relevant capabilities. Finally, the qualitative evaluation should be redesigned as well. The interviews have shown that focusing on the service provider as a unit of analysis is not something that is seen as beneficial and will probably be the responsibility of the service provider to directly engage with the participants in order to improve their own services. Qualitative evaluation can still add value though, but it should focus upon the system and processes of the SIB as a whole, by speaking with all involved stakeholders.

Redesign of the evaluation tool

With the learnings from looking at the existing Social Impact Bonds and interviewing stakeholders that are actively involved in SIBs, the evaluation tool should be reconsidered where that is deemed necessary. Based on the conclusions of the previous chapter, this chapter will provide the characteristics of the evaluation tool as it is proposed based on the empirical input. This chapter will therefore be built up similarly to chapter 5. To start off, the evaluation part of the tool will be discussed. Starting with the evaluation of the outcomes, then the evaluation of the well-being of the participants and finally the qualitative evaluation. Second, the monitoring will be discussed. Leading to an overview of the final evaluation tool.

7.1. Evaluation

The evaluation is still the most important part of the tool and will look back at the outcomes of the SIB, assessing the impact it has had and evaluating the processes. These three aspects will all be discussed in this section. The evaluation of the outcomes will remain very similar to initially designed, while the other two aspects will be redesigned. For the evaluation of the impact on participants, a new quantitative questionnaire will be designed based on the conclusions of the previous chapter. Also the qualitative evaluation will now be changed and halved, based on the interviews. While the evaluation of the processes of the SIB remains, the evaluation of the processes of the service provider is scrapped. This is seen as adding little value to the evaluation and to shorten the evaluation tool, increasing the achievability, that specific aspect is therefore not included in the final design.

7.1.1. Evaluating outcomes

As said, this part will remain practically the same to the initial design. That design recognised that outcomes will be different for all SIBs, as outcome indicators will be set in alignment with the goals of the individual SIB. Nonetheless, the criteria for those indicators can be set beforehand, to ensure that appropriate indicators are being used and evaluation can take place properly. It is restated by the interviewed stakeholders that it is important that the indicators that are being used are agreed upon and objective, apolitical. This will be therefore be maintained in the design. Evaluating outcomes should follow the same four steps:

1. *Determine criteria for baseline data* While a sizeable group of SIBs have been using a control group, based on the conclusions of the previous chapter it is still advised to use baseline data for the evaluation of the outcomes. While the criteria that are used to collect the baseline data should make sure that the baseline represents the participants as much as possible, it is also clear that due to regulations or sometimes just general poor data collection, it won't always be possible to do exactly that. This should either be assumed through a negotiation process as initially suggested, but potentially the most accurate solution would be to let the evaluator do research on those factors and propose a controlling method for those based on existing data and/or literature.

2. *Collect data at start of the intervention for baseline and participating service users* Nothing has changed to this step of the evaluation. First the starting position of the participants should be collected, but then the baseline should also be collected for that specific group of participants based on

their characteristics. The baseline should be as current as possible, but it is recognised that sometimes the baseline will be set based on historical data.

3. *Collect data after the intervention has completed for baseline and participating service users* The same process as in step 2 will take place but then directly after the intervention. When the baseline data can also be reassessed this should be done to be able to control better for exogenous effects.

4. *Analyse data to determine final outcomes* When baseline data could be collected before and after the intervention, the analysis will be different than when baseline data could be collected only once. In the first instance, the achieved outcomes should be compared with the baseline and subsequently corrected for the change in the baseline to correct for the exogenous effects that have taken place during the intervention. If not, the outcomes of the participants should simply be compared to the baseline.

The evaluation of the outcomes will purely look at the intended outcomes for which outcomes are constructed, although the same approach can also be used for broader foreseen outcomes. When possible it will also look at the exogenous effects of the SIB system though, but this will mostly be possible for indicators allowed for the collection of baseline data before and after the intervention. It can obviously be the case that baseline data can be collected before and after the intervention for certain indicators, but not for others. In those circumstances the correction for the exogenous effects should only take place for those indicators where they can be established.

7.1.2. Evaluating impact on participants

In the analysis chapter, the capabilities that are most relevant for SIBs are identified. The original questionnaire as proposed by Anand (2007) is quite extensive and as the practical achievability for the evaluation also needs to be taken into account, it might be wise to make the questionnaire more specific. This will make the questionnaire better to handle for the evaluator as well as for the unit of observation, in this case the participants. Therefore a questionnaire will be developed that reflects the main capabilities that are targeted in SIBs, but also reflect the criteria that are identified earlier in this thesis on the evaluation tool. To do so, information is gathered on the questions and answering methods that can be identified for the relevant capabilities from earlier studies. Later, this is used as a basis to create a new questionnaire that is shorter than the initial one, but still appropriately evaluates the relevant capabilities and can evaluate the impact SIBs have on the well-being of service users.

There are three studies by Anand et al available, in which different types of questions are used based on the area of questioning [6, 7], but also one which includes the perception of the service users on others [5]. Even though perception on others is an interesting aspect to asking questions in general, but it could also impose a significant bias on the interviewees answers, based on their perspective of their surroundings. Especially when comparing the different SIBs, this might not be a good enough method of comparison. When making use of the BHPS survey however, Anand et al, also provided a list of questions categorized by the capabilities, which provides useful insights and gives six questions on Bodily Health, two questions on Bodily Integrity, one measurement scale of Senses, Imagination and Thought, three questions on Practical Reason, eight questions on Affiliation and twelve questions on Control over one's Environment [6]. The third research consists of 65 non-categorised questions [7].

Brandolini and Alession [12], performed a survey on the Quality of Life, which hasn't categorised to exactly the same capabilities as the central capabilities, but has questions grouped on the topics of Health, Education, Employment, Housing and Social Relationships. Another example is that of Clarke, that asks a set of twelve questions on a four point scale on their distress due to unemployment [19].

These different studies are used as a basis to identify the different questions and are adapted in order to reflect the differences in interviewees age, perspective and situation that can be encountered when looking at the different SIBs.

Examples of questions per capability

Bodily Health The first category to identify questions for is Bodily Health. When looking at the SIBs, there have been various SIBs related to Bodily Health, from reducing homelessness to improving long-term health outcomes. Looking at the mentioned studies and their health related questions, the following questions have been used:

1. Does your health in any way limit your daily activities compared to most people of your age? (Y or N) [6]
2. Do you eat meat, chicken, fish every second day (Y or N) [6]

3. Would you like to be able to eat meat, chicken, fish at least every second day, but must do without because you cannot afford it? (Y or N) [6]
4. If you could choose, would you stay here in your present home or would you prefer to move somewhere else? (Y or N) [6]
5. (Even though you may not want to move) Do you expect to move in the coming year? (Y or N) [6]
6. I feel the scope for intellectual stimulation in my life is: (score 1-7, very good to very inadequate) [5]
7. I live a healthy life, for my age (score 1-7, very good to very inadequate) [5]
8. Does your health in any way limit your daily activities compared to most people of your age? (Y or N) [7]
9. Are you able to have children (Y or N) [7]
10. Do you eat fresh meat, chicken or fish at least twice a week (Y or N) [7]
11. Is your current accommodation adequate or inadequate for your current needs (Adequate or Inadequate) [7]
12. Are you prevented from moving home for any reason [7]
13. In general terms, how would you describe your health (1-5) [12]
14. Do you suffer from any chronic illnesses (Y or N) [12]
15. Do you suffer from any form of disability (Y or N) [12]

There is only one question that overlaps between the different studies. Looking at the purpose of the evaluation framework, the questions around suffering from chronic illnesses or disability do not appear to be appropriate as that can not be expected to change over the course of a SIB, the same can be said about the ability to have children, which also can be seen as a sensitive question [84]. The questions around being able to move home or accommodation, can also be grouped, therefore questions 2,4,8 and 13 can be singled out for this capability.

Bodily Integrity The diversity of available questions for the Bodily Integrity are slightly less diverse as the concept is more narrow and therefore questions can be asked on defined topic areas around experiencing violent behaviour and being able to move freely from place to place [47]. The questions that have been used previously:

1. Do you normally have access to a car or van that you can use whenever you want to? (Y or N) [6]
2. Does your accommodation have any of the following problems? Vandalism or crime in the area? (Y or N) [6]
3. I feel the scope to act with personal integrity in my life is: (score 1-7, very good to very inadequate) [5]
4. Please indicate how safe you feel walking alone in the area near your home during daytime (score 1-7) [7]
5. Have you ever been the victim of some other form of violent assault or attack? (Y or N) [7]
6. How likely do you think it is that you will be a victim of violent assault or attack in the future (score 1-7) [7]
7. Have you ever been a victim of sexual assault (Y or N) [7]
8. Please indicate how vulnerable you feel to sexual assault or attack (score 1-7) [7]

9. Have you ever been a victim of domestic violence (Y or N) [7]
10. Please indicate how vulnerable you feel to domestic violence in the future (score 1-7) [7]

Questions are predominantly around experiencing violent behaviour and being a victim of it. However, some of the SIBs are also engaged in reducing incarceration and thus reducing criminal activity. Therefore a question that looks at that other side could be *Have you recently committed violent, sexual or domestic assault or attack*, although that definitely will be seen as a sensitive question, for which answers should be carefully monitored [84]. For the purpose of the SIB, there is no real value in separating the different types of assault or violence, therefore they can likely be grouped. Asking about the existing experiences is likely to be static, but can still be relevant to find whether that is correlated with their perspective of the future, but also to identify the problems of a group. The answers to the likelihood of experiencing violence in the future is especially important because it can related back to improving this capability in different forms. For instance, by moving to a more safe environment, but also if one would receive training in managing potential dangerous situations, this could the reduce the likelihood of the person experiencing violence, and therefore have an effect on the capabilities of that person. Finally the question around being able to use a car/van is interesting, but there can be issues around validity of that question, as this can also just be due to economic resources, while having a car or van are clearly not the only potential methods of moving freely from place to place. A question around being able to move freely from place to place can however be asked in a more general sense: *On any given day, could you go wherever you want to go (Yes, No because I cannot afford it, No because of other reasons.*

Senses, Imagination and Thought In several SIBs, the capability of Senses, Imagination and Thought plays an important role, mainly through educational attainment. However, questions on this topic also investigate creativity and imagination in more depth:

1. What is the highest educational qualification that you have achieved (With answers that consist A-levels (and equivalent) or higher being coded as Yes, the rest as No) [6]
2. I feel the scope for intellectual stimulation in my life is? (score 1-7, very good to very inadequate) [5]
3. What is the highest educational or work related qualification you have (With answers that are A-levels (and equivalent) or higher being coded as Yes, the rest as No) [7]
4. How often do you use your imagination and or reasoning in your day to day life [7]
5. I am free to express my political views [7]
6. I am free to practice my religion as I want to [7]
7. Give highest qualification earned [None, Elementary school, Middle school, Professional secondary school diploma (3 years of study), High School, Associate's degree or other short course university degree, Bachelor's degree, Post-graduate qualification] [12]

In three of the studies, this section is being determined by a single question and in two of those, the capability is determined by the highest educational attainment [6, 12]. This question also appears in the fourth example and is clearly a core part of measuring this capability. Most SIBs however are also focusing on supporting young people to achieve their secondary school attainment, to improve their outlook on life and increase chances of them continuing in education (For instance in the workforce development SIBs). Therefore, a question should be taken into account that asks *What is the highest educational qualification that, if you had the time and money resources to study, you think you could be able to achieve?*, to test whether their perception of future education options changes and their capabilities have increased. This can combined by the religious and political freedoms to completely measure this capability,

Affiliation The Affiliation capability can be split into two areas that are touched upon in the SIBs, Employment and social relationships. In the various studies, these are also the topics that are mostly questioned in the questionnaires:

1. Do you and/or your household pay for a week's annual holiday away from home? (Y or N) [6]
2. Would you like to be able to pay for a week's annual holiday away from home but must do without because you cannot afford it? (Y or N) [6]
3. Do you and/or your household buy new, rather than second hand, clothes (Y or N) [6]
4. Would you like to be able to buy new, rather than second hand, clothes, but must do without because you cannot afford it? (Y or N) [6]
5. Do you and/or your household have friends or family for a drink or meal at least once a month [6]
6. Would you like to be able to have friends or family for a drink or meal once a month, but must do without because you cannot afford it? (Y or N) [6]
7. Have you recently been thinking of yourself as a worthless person? (Not at all to much more than usual) [6]
8. Have you recently felt that you were playing a useful part in things (Not at all to much more than usual) [6]
9. I feel the scope to form satisfying social relations in my life is (score 1-7, very good to very inadequate) [5]
10. I respect, value and appreciate other people (score 1-7) [7]
11. Do you normally have at least one week's (seven days) annual holiday away from home? (Y or N) [7]
12. Do you normally meet up with friends or family for a drink or meal at least once a month (Y or N) [7]
13. Do you tend to find it easy or difficult to imagine the situation of other people (i.e. to put yourself in others shoes) (score 1-7) [7]
14. Have you recently been thinking of yourself as a worthless person (score 1-4) [7]
15. Outside of any employment or work situation, have you ever experienced discrimination because of your [race, sexual orientation, gender, religion, age]? (Y or N) [7]
16. Outside of any employment or work situation, do you think that in the future you will be discriminated against because of your [race, sexual orientation, gender, religion or age]? (Y or N) [7]
17. Have you ever lost a job at any time and remained unemployed for at least 6 consecutive months (N.B. not when a first-job seeker) (Y or N) [12]
18. Have you ever been on wage supplementation at 0 hours for at least 6 consecutive months? (Y or N) [12]
19. When you were unemployed or on wage supplementation, as compared with when you were working, how would you rate your [role in the family, health, ability to have relationships with other people, your confidence in yourself, your ability to exploit your free time, overall apart from the economic aspect you feel/felt yourself to be]? (score 1-5) [12]
20. For all employed members, how do you judge your work as regards to [environmental conditions (physical and social), dangerousness for life or health, demandingness, interestingness, consideration by others, concern about losing your employment, overall satisfaction apart from the economic aspect] [12]

21. In the previous year, did you do anything to find employment (temporary or otherwise) or to change your employment (Y or N) [12]
22. Can you tell me why you did not look for employment? (potential answers are: Family reasons, household's income was sufficient, it would not have been worthwhile economically, because of the difficulty of finding work, for health/disability reasons, I was waiting for public competitive exams, I was studying, I was doing/waiting to do military service, other) [12]
23. Were your parents still alive at the end of last year? (Y or N) [12]
24. Did you have brothers or sisters who were not residing with you at the end of last year (Y or N, If yes how many brothers, how many sisters) [12]
25. Did you have children who did not reside with you at the end of last year (Y or N, If yes how many children) [12]
26. What is your labour market status (Unemployed, Employee, Self-employed) [19]

Because the topics employment and especially social relationships are so broad, there are a lot of exemplary questions in measuring this capability. Some of the questions, in two different studies, go into the topics of being able to go on holiday or host nights with friends/family [6, 7]. This in some cases measures more the economic resources of a household than employability or social relationships. Although employment can potentially be linked with economic resources (obviously depending on issues such as the local welfare system), it should be seen as a separate thing. Therefore, the questions about annual holiday/second hand clothes hold less relevance for the SIBs, but the question around having family or friends over at least once, can play into having a safe home environment, as well as having the right relationships and therefore can measure this capability.

There are questions surrounding self-respect, forming social relationships and respecting other people, which can be seen as important cornerstones in developing yourself towards a better life and should therefore be used in a general survey. Finally, questions on employment and discrimination should be bundled in single questions, but also here something on the actual capabilities is missing and should be added: *How confident are you that if you currently were to look for a (new) job, that you would be able to find a job (score 1-7)*. Finally, the questions around family measure the presence of family but not the engagement with, what for instance plays a role into several SIBs concerning foster care. Also, the presence of siblings, shouldn't significantly influence the capabilities of a person. Therefore, questions around this section should be rephrased into a new set of questions that tests whether someone is caring for another person, is residing with people who are caring for him/her and whether they have regular contacts with family contacts. These three topics could be tested by the following questions:

- Do you have parents or children whom you see less than once a month? (Y or N)
- Are you currently caring for a person who is residing with you (children, or other)? (Y or N)
- Are there any people who you would want to be caring for, but cannot care for them for any reason? (Y or N)
- Are you currently residing with someone who is caring for you? (Y or N)
- Are there any people that you would like to have caring for you, but you are not residing with for any reason? (Y or N)

Practical Reason The capability of Practical Reason was related as a secondary capability to some SIBs based on being able to form a conception of the good. General questions however are focused on mental health:

1. Have you recently felt capable of making decisions about things (Much less capable than usual to More so than usual) [6]
2. Have you recently felt you couldn't overcome your difficulties (Not at all to Much more than usual) [6]

3. Have you recently been able to face up to problems (Much less than usual to More so than usual) [6]
4. My idea of a good life is based on my own judgment (score 1-7) [7]
5. I have a clear plan of how I would like to life to be (score 1-7) [7]
6. How often, if at all, do you evaluate how you lead your life and where you are going in life (score 1-7) [7]
7. Outside of work, have you recently felt that you were playing a useful part in things? [7]

The questions on Practical Reason can see some threats to validity due to concepts being vague (i.e. making decisions about things), but these questions can still give a good sense around decision making capability. Therefore, the questions will be slightly rephrased into more clear concepts. Questions 5 and 6, can be used directly, while question 1 will be rephrased to *Have you recently felt capable of making decisions about crucial parts of your life.*

Control over one's Environment The other, more secondary, capability that can sometimes be related to SIBs is that over the Control over one's Environment. Not all studies investigate this, but the generated questions are:

1. Did you vote in the most recent general election (Y or N, those who couldn't vote are coded one) [6]
2. Does your health keep you from doing some types of work? (Y or N) [6]
3. For work you can do, how much does your health limit the amount of work you can do (Can do nothing to No) [6]
4. How dissatisfied or satisfied are you with your [life overall, health, income of your household, house/flat, husband/wife/partner, job, the amount of leisure time you have, the way you spend your leisure]? (score 1-7, Not satisfied at all to Completely satisfied) [6]
5. I am able to participate in the political activities that affect my life if I want to (Y or N) [7]
6. For which of the following reasons, if any have you not bought your home? (Forced not for reasons, affordability, difficulty obtaining mortgage, chose not to buy for other reasons, am a home owner) [7]
7. When seeking work in the past, have you ever experienced discrimination because of your [race, sexual orientation, gender, religion, age]? (Y or N) [7]
8. When seeking work in the future, how likely do you think it is that you will experience discrimination because of your [race, sexual orientation, gender, religion, age]? (Score 1-7) [7]
9. How likely do you think it is that within the next 12 months you will be stopped and searched by the police when it is not warranted? (Score 1-7) [7]
10. To what extent does your work make use of your skills and talents? (Score 1-7) [7]
11. At work, have you recently felt that you were playing a useful part in things? (Score 1-4) [7]
12. Do you tend to find it easy or difficult to relate to your colleagues at work? (Score 1-7) [7]
13. At work, are you treated with respect (Score 1-7) [7]

Questions on politics should include the option to say that someone is not capable to engage politically due to age restrictions. Although this can still be seen as a general restriction on political freedom, these restrictions are placed upon almost every country and would skew results when the questionnaire is taken by under aged interviewees. Some questions can potentially be better related to other capabilities and will be treated as such. Therefore, questions 2 to 6 will be included in the Bodily Health capability, questions 7 and 8 to the Affiliation capability and question 9 to the Bodily Integrity capability. This leaves questions 1 and 9 that are relevant, while the other questions are more related to social relationships than actual control over one's environment.

Building up the overall questionnaire

As the SIBs target very diverse people, the questionnaire should be able to be relevant for all of them and questions should be phrased in clear, concise wording so that all potential interviewees will be able to engage with the questions and can answer them clearly. Concepts should be made clear to increase validity and the questionnaire as a whole should be concise to enable interviewees to complete the whole questionnaire with relatively little effort in time and concentration. Because some questions on capabilities can be perceived as sensitive questions, the questionnaire should make sure that interviewees are relatively comfortable answering the questions to increase the reliability of their answers. To secure this, interviewees should complete the questionnaire without the interviewer knowing the answers (a self-administered test) and anonymously [84].

The full questionnaire will consist of an introduction to introduce the study and purpose of the questionnaire, but will also include general questions on age, gender and country of residence. These three items can be seen as relatively common and should be used to monitor the answers of some questions as they can depend on age, gender or country of residence (i.e. different countries can also have different education systems). The full questionnaire will therefore be presented in this section, based on the mentioned requirements and the questions as selected or defined in previous sections of this chapter.

Introduction: Thank you for agreeing to take part in this survey. We are trying to get a better understanding how the programme of [Service Provider] has effect on you as a person and we will use this survey to gain your thoughts and insights on your perspective on certain aspects of your life. All answers will be collected anonymously and will be kept in the strictest confidentiality. In total the survey should take approximately 10 minutes to complete.

General questions:

- What is your month and year of birth?
 - Month:
 - Year:
- What is your gender?
 - Male
 - Female
 - Other (i.e. Transgender)
 - Prefer not to say
- What is your current country of residence?

Questions related to the person's capabilities:

- Do you eat meat, chicken or fish at least twice a week?
 - Yes
 - No
- If you could choose, would you stay here in your present hoe or would you prefer to move somewhere else?
 - Yes
 - No
- Does your health in any way limit your daily activities compared to most people of your age?
 - Yes
 - No
- Does your health keep you from doing some types of work?
 - Yes

- No
 - Prefer not to Say
- In general terms, how would you describe your health?
 1. Not healthy at all
 2. Slightly unhealthy
 3. Adequately healthy
 4. Above average healthy
 5. Very healthy
- Do you feel that on any given day, you could go wherever you want?
 - Yes
 - No, because I cannot afford it
 - No, for other reasons
- Does your accommodation have vandalism or crime in the area?
 - Yes
 - No
- Have you ever been the victim of violent or sexual assault, domestic violence, or any other form of attack?
 - Yes
 - No
 - Prefer not to say
- How likely do you think it is that you will be a victim of violent or sexual assault, domestic violence, or any other form of attack in the future?
 - 1 - Very unlikely, to 7 - Very likely
- How likely do you think it is that within the next 12 months you will be stopped and searched by the police when it is not warranted
 - 1 - Very unlikely, to 7 - very likely
- What is the highest educational qualification that you have achieved until now?
 - None
 - Elementary school
 - Secondary school diploma (including GCSE results)
 - Sixth Form (including A-levels)
 - Further education other than University
 - Bachelor's degree from a university
 - Post-graduate qualifications
 - Prefer not to say
- What is the highest educational qualification that, if you had sufficient time and money, you think you could be able to achieve?
 - None
 - Elementary school
 - Secondary school diploma (including GCSE results)

- Sixth Form (including A-levels)
 - Further education other than University
 - Bachelor's degree from a university
 - Post-graduate qualifications
 - Prefer not to say
- How likely do you think it is that you will achieve that educational qualification?
 - 1 - Very unlikely, to 7 - Very likely
- Do you feel free to express your political views?
 - Yes
 - No
 - Prefer not to Say
- Do you feel free to practice your religion the way you would want to?
 - Yes
 - No
 - Prefer not to Say
- Do you and/or your household have friends or family for a drink or meal a least once a month?
 - Yes
 - No
 - Prefer not to Say
- Would you like to be able to have friends or family for a drink or meal at least once a month, but must do without because you cannot afford it?
 - Yes
 - No
 - Prefer not to Say
- Have you recently been thinking of yourself as a worthless person?
 1. Not at all
 2. Less than usual
 3. Slightly more than usual
 4. Much more than usual
- To what extent do you feel that you respect, value and appreciate other people, independent of their race, sex or religion?
 - 1 - Not at all, to 7 - I fully respect other people
- Are you currently employed?
 - Yes
 - No
 - Prefer not to Say
- Have you ever lost a job at any time and remained unemployed for at least 6 consecutive months? (not when you were a first-job seeker)
 - Yes

- No
 - Prefer not to Say
- If you are currently unemployed, would you like to find a job in the next twelve months?
 - Yes
 - No
 - Prefer not to Say
- If you are currently employed, would you like to switch jobs in the twelve months?
 - Yes
 - No
 - Prefer not to Say
- How likely do you think it is that, if you currently were to look for a (new) job, that you would be able to find a job?
 - 1 - Very unlikely, to 7 - Very likely
- When seeking work in the future, how likely do you think it is that you will experience discrimination because of your:
 - Race (1 - Very unlikely, to 7 - Very likely)
 - Sexual orientation (1 - Very unlikely, to 7 - Very likely)
 - Gender (1 - Very unlikely, to 7 - Very likely)
 - Religion (1 - Very unlikely, to 7 - Very likely)
 - Age (1 - Very unlikely, to 7 - Very likely)
- Outside of any work or employment situation, how likely do you think it is that in the future you will experience discrimination because of your:
 - Race (1 - Very unlikely, to 7 - Very likely)
 - Sexual orientation (1 - Very unlikely, to 7 - Very likely)
 - Gender (1 - Very unlikely, to 7 - Very likely)
 - Religion (1 - Very unlikely, to 7 - Very likely)
 - Age (1 - Very unlikely, to 7 - Very likely)
- Do you have parents or children whom you see less than once a month?
 - Yes
 - No
 - Prefer not to Say
- Are you currently caring for a person who is residing with you (children, parents, or other)?
 - Yes
 - No
 - Prefer not to Say
- Are there any people who you would want to be caring for, but cannot care for them for any reason?
 - Yes
 - No
 - Prefer not to Say

- Are you currently residing with someone who is caring for you?
 - Yes
 - No
 - Prefer not to Say
- Are there any people that you would like to have caring for you, but you are not residing with for any reason?
 - Yes
 - No
 - Prefer not to Say
- Have you recently felt capable of making decisions about crucial parts of your life?
 1. Much less capable than usual
 2. Less capable than usual
 3. More capable than usual
 4. Much more capable than usual
- I have a clear plan of how I would like life to be?
 - 1 - Not at all, to 7 - I know very clearly how I would like life to be
- Did you vote in the most recent general election?
 - Yes
 - I was not able to vote due to my age
 - I was not able to vote due to other reasons
 - I did not vote
- For which of the following reasons, if any, have you not bought your home?
 - Forced, not for particular reasons
 - Affordability
 - Difficulty obtaining mortgage
 - Chose not to buy for other reasons
 - I am a home owner

Closing: Many thanks for completing the questionnaire. Please use a provided envelope to include your answers in or click on finish, in case you answered the electronic questionnaire. Your time and answers are much appreciated and will be processed anonymously and with the strictest confidentiality.

7.1.3. Qualitative evaluation

Initially the qualitative evaluation would look at the service provider and analyse the processes of the intervention itself. However, based on the input of the interviewees, this would not add significant benefit to the evaluation of the SIB. However, evaluating the processes of the SIB as a whole by interviewing the stakeholders does in fact add value and provide valuable insights into the processes of the SIB system and the interactions between the stakeholders. The interview should be qualitatively to gain deeper insights and allow for additional deeper questions during the interview in order to create a full picture of the SIB and its processes. All stakeholders that were directly involved in the SIB should be interviewed and those interviews should be analysed to gain more insights and the processing of that analysis should provide learnings for all stakeholders towards future SIBs.

Therefore a new questionnaire is designed with the following questions:

1. Why did you participate in this Social Impact Bond?

2. As an organisation, have you benefited from the SIB?
3. Do you think the overall goals of the SIB have been realised?
 - If not, why?
4. Do you think your individual goals for this SIB have been realised?
 - If not, why?
5. How could you have benefited more from the SIB?
6. How have the relationships with the other stakeholders developed?
7. How could the ... have benefited more from the SIB?
 - Commissioner
 - Investor
 - Service Provider
 - Participant
8. How would you have made the agreements with the other stakeholders differently than set out in this SIB, using what you have learned in this SIB?
9. What outcomes were the result of this SIB that you hadn't foreseen at the start?
10. In hindsight, what changes would you have made to the indicators that have been used for this SIB?
11. Overall, could you explain which parts of the SIB could be improved and why?

These generic questions should give a broader insight into the process of the SIB and can be used to provide learnings for the future of SIBs. For question 7, the questions should be asked for all of the stakeholders apart from the stakeholder that is being interviewed at that moment. This is to gain more insights in the perceptions of all stakeholders into the others of the SIB.

7.2. Monitoring

Based on the analysis and inputs from the interviews, it is concluded that monitoring should focus purely on the monitoring of inputs, outputs and outcomes. The initial proposal of monitoring the process of the service provider does not add significant value to the SIB and is not something that should be added to the evaluation tool because of that. Also, keeping track of the resources of the service provider is not needed as that is specified in the agreements and monitoring that doesn't add value or allow for change anyway. It should still be made explicit what information should be included in the monitoring though. The monitoring should take place every three months and should include the current state of the intervention, based on the inputs in number of participants, the status of those participants and the outcomes that have been achieved. Based on the baseline data that has been collected before the start of the intervention, it should be possible to make a prediction based on the current number of achieved outcomes, what the final result would be. This prediction should be adjusted at each monitoring moment and can provide additional insights into the progress of a SIB in a fair way. The one financial aspect that should be included in the monitoring is the value of the achieved outcomes until that point in time and the expected financial return based on the projection of outcomes. The information that therefore should be collected and discussed would therefore be:

- All participants that are currently participating in the intervention
- When relevant: the stage current participants are in and what part of the intervention they are involved in
- A comparison of the current number of participants and the expected number of participants until that point in time

- Projection of expected outcomes based on the number of participants, the stage they are in and the expected success rates
- A comparison of the total expected outcomes as currently standing and the negotiated expected outcomes as set when the SIB started
- A financial overview of the expected return, based on the achieved outcomes and the projection of expected outcomes

7.3. Governance

While the interviews and the analysis has provided additional insights and a further redesign of the evaluation tool, there are no significant changes to the governance of the tool. The importance of getting access to the right data is emphasised [24, 89] and the necessity of having frequent monitoring has been shown. In multiple interviews there was an emphasis on trust between the stakeholders and making sure that each one is allowed to do the things it is good at. This will therefore also definitely go for the evaluator that will be hired to perform the evaluation, but still transparency of the evaluation process and results is required. In general however, the governance of the tool will be similar to that described in chapter 5.

7.4. Conclusions of redesign

This chapter has proposed a redesign of the initial evaluation tool. The tool has been adjusted based on the input as delivered by several interviewed stakeholders and the analysis of the existing SIBs. This has led to some significant changes in the tool, but also affirmed the importance of several items. The result is an evaluation tool that looks at both evaluation and monitoring, evaluating all aspects of the system of a SIB and focusing on both the formative and normative functions of evaluation.

Evaluation will take place based on data that is collected before and after the intervention. To increase the validity of the evaluation and evaluate the actual added value of the intervention, baseline data is collected for all indicators that are being used. When possible this baseline is also set before and after the intervention in order to also control for exogenous effects, although it is concluded that under no circumstances it will be possible to fully control for exogenous effects. The impact of the intervention on the participants will also be evaluated by using a questionnaire that the participants will fill in before and after the intervention. This questionnaire, based on the Capabilities Approach, is based on the capabilities that were identified as the core priorities during the analysis of the existing SIBs. Finally a qualitative evaluation is performed with all direct stakeholders of the SIB to gain deeper insights in the processes of the whole SIB and learn lessons that can be implemented in future SIBs. This all is supported by frequent, ideally quarterly, monitoring. The main inputs, outputs and outcomes are assessed every three months and should be discussed amongst all stakeholders to update all stakeholders, but also allow for slight adjustments for the process whenever that is both necessary and also possible.

In terms of time, FTE or costs related to this evaluation tool, a conscious decision has been made to not attach those aspects to this design for a couple of reasons. First, it is simply impossible to estimate the full costs of evaluation as it depends on each individual SIB what data is being collected and for how many users. Secondly, costs can put a limit on the validity of the evaluation and should never be the starting point of setting up evaluation. The starting point should always be setting up an evaluation that has the highest validity possible and falls within the set requirements of the thesis. It is clear, however, that this evaluation will take up a significant proportion of the costs of a SIB.

Conclusions

This thesis has gone through a design cycle with the objective *to design and test an evaluation tool that evaluates Social Impact Bonds*. To satisfy that objective, first literature has been used to gain more insights in evaluation tools and evaluating social impact, to set out the criteria for an evaluation tool of Social Impact Bonds. After that, the workings of SIBs have been set out more, and research is done to the Capabilities Approach in order to evaluate the impact that SIBs have on the participants of SIBs. This has led to an initial design of an evaluation tool for Social Impact Bonds. Following the design phase, the design was tested by interviewing several stakeholders of SIBs and analysing all existing SIBs that have been launched so far. These tests have led to a redesign of the evaluation tool and inspired a final proposal for a generic evaluation tool for Social Impact Bonds. The outcomes and conclusions will be discussed in more detail by answering the research questions that have been set out in section 1.2 of this thesis.

1. Which set of criteria should be met to develop a generic evaluation tool for Social Impact Bonds?

This first research question has been answered in chapter two of this thesis. Based on existing literature, the different aspects of evaluation have been investigated and have led to a set of criteria that should be used to develop the evaluation tool. In this development, it is discussed who the evaluation should perform, when it should take place, the aspects that influence the validity of the tool and practical achievability of the tool. A set of seven criteria were developed:

- The evaluation should be both quantitative and qualitative
- The evaluation should have separate evaluations that have service users and service providers as units of analysis
- Outcome indicators should focus on objective and agreed upon criteria
- Monetisation should not be part of the evaluation itself
- Effects on society and on people's well-being should be evaluated
- The evaluation should use baseline data and ex-ante, ex-post evaluation
- Monitoring and subsequent discussion of the monitoring should take place at least every three months and in between each cohort of a project (when applicable)
- The evaluation should be performed by a third party

These criteria are such that it should lead to an evaluation that is valid, but also cost-efficient and can provide learnings for future SIBs. These criteria have formed the basis for the design of the eventual evaluation tool. One aspect remains troublesome though, the causality of the tool. While the criteria have been developed such that the validity in all ways is as high as possible, it is recognised both in

setting up the criteria as later in testing the tool, that fully proving causality will be impossible. It is simply not possible to exclude all exogenous effects in an open society where people engage in all aspects of life. Nevertheless, the criteria of using baseline data, looking at unintended outcomes and evaluating the broader effects on the participant's well-being will ensure that grounded comments can be made on causality.

2. What are the characteristics of the designed evaluation tool for Social Impact Bonds?

Based on the criteria as developed from the first research question, and the more substantial information on SIBs as well as the Capabilities Approach, the characteristics of the evaluation tool specifically for SIBs are defined. This has led to an initial design of the tool that covers both criteria of the evaluation tool and the system of SIBs. In general the design can be split over an evaluation and a monitoring part, which both consist of multiple aspects.

- *Evaluation*

The evaluation part of the tool will look backwards at the SIB and evaluate the outcomes, the impact and the processes of the SIB. First of all, the outcomes are evaluated using baseline data to see the actual improvement of the outcomes indicators compared to the baseline. Data for the outcome indicators will be collected before and after the intervention. When the indicators allows for it, baseline data will also be collected before and after the intervention. If not, baseline data will be evaluated only before the intervention takes place. The eventual outcomes can be found by comparing the change in the outcome indicator for a participant, with the baseline data. When baseline data was collected also after the intervention, this also makes it possible to control for exogenous effects such as a financial crisis.

It is also seen as important that the perspectives of the participants is taken into account for the evaluation. These are the people who have received the services. In many cases we might think that those services have had a positive impact on their lives, but it is crucial to also evaluate that from their perspective. Using the Capabilities Approach, the change in well-being of the participants is evaluated by giving a questionnaire before and after the intervention. With the Capabilities Approach, it is not only possible to evaluate the impact on well-being, but also to look at specific capabilities and identify areas where potentially unintended outcomes have been realised, either positive or negative.

Finally, a qualitative evaluation is designed where the service user is the unit of observation, but now the service provider is the unit of analysis. By using qualitative interviews, deeper insights can be extracted on the delivered intervention to see if and how that intervention can be improved. By using the participants as the unit of observation, this is tested by the people who are directly engaging with the intervention.

- *Monitoring*

Monitoring is split into two aspects. First, the progress of the interventions is monitored through the hard data on inputs, outputs and outcomes. This to give all stakeholders insight in the status of de SIB. As set out by the criteria, monitoring will take place at least every three months. By doing so it should spur adaptive decision making based on the insights gained by these frequent monitoring results. To correctly monitor all inputs, outputs and outcomes of a SIB, at least the following items should be considered:

- All human capital that has been involved in the execution of the intervention (this includes both human capital that was foreseen and/or is being paid for as additional human capital that has not been paid for or is involved on a voluntary basis)
- All financial capital that has been used to date
- All other resources that have been used for the intervention
- A comparison of the used resources and the budgeted resources for the intervention until that point in time
- All participants that are currently participating in the intervention

- When relevant: the stage current participants are in and what part of the intervention they are involved in
- A comparison of the number of participants and the expected number of participants until that point in time
- Projection of expected outcomes based on the number of participants, the stage they are in and the expected success rates
- A comparison of the total expected outcomes as currently standing and the negotiated expected outcomes as set when the SIB started

Additionally, the qualitative aspect of the evaluation should also be performed on a sample of the participants to track the processes of the intervention and use those insights to adjust the intervention whenever that is needed.

3. What are the results of testing the evaluation tool on Social Impact Bonds?

The developed evaluation tool is then tested in two ways. First, several stakeholders of existing SIBs have been interviewed on their perspectives on evaluating Social Impact Bonds and their views on the designed evaluation tool. Secondly, all existing SIBs have been analysed on their design, goals and evaluation methodologies. The results of those two exercises have led to some conclusions on the criteria of evaluation and the designed evaluation tool.

Based on the empirical analysis, it was confirmed that it was important that an independent evaluator performed both the evaluation and monitoring aspects of the evaluation tool. Also, the approach of using baseline data appeared to be in line with the majority of the SIBs and was seen as a preferred methodology by the stakeholders. In general there were worries around the practical achievability of the evaluation tool though. SIBs are seen as an expensive product and therefore the evaluation tool should be designed as efficiently as possible. Not all aspects of the initial design were seen as adding value to the SIB as a whole or it was expected that when there were no direct incentives for stakeholders to engage in certain aspects of evaluation, that it then would be unlikely that costs would be allocated for it.

As a result, the evaluation of the impact on the well-being of participants have been redesigned. Based on the analysis of the existing SIBs, the six capabilities that were actively targeted in the SIBs were identified and it was concluded that the questionnaire should be redesigned based on those six capabilities, to shorten the questionnaire, lower the costs, but also increase validity of the tool.

The tests also concluded that evaluating the resources and processes of the service provider had little added value to the SIB, contrary to what originally was thought. However, the interviews showed that it was thought that there was added value in evaluating the processes of the SIBs as a whole. Getting deeper insights in the processes of a SIB will provide valuable learnings for the future for all stakeholders.

Finally, in the criteria of the tool it was stated that monitoring should take place at least every three months. There were different opinions on the frequency, but they were in line with this frequency. However, especially the service providers weren't directly satisfied with a much higher frequency. Monthly monitoring meetings is seen as somewhat too much, therefore quarterly appears much more sensible for all parties.

4. How should a generic evaluation tool be redesigned based on the case study on Social Impact Bonds?

The results of testing the evaluation tool has provided some valuable lessons for the design of the evaluation tool. Therefore, the tool has been redesigned based on those lessons and made more cost-efficient, while maintaining or improving the validity of certain aspects. An overview of the final tool can be found in Appendix A.

While the evaluation of the outcomes have remained the same, the questionnaire for evaluating the impact on the well-being of the participants has been redesigned based on the six identified capabilities that have shown to be most important. This redesign has reduced the length of the questionnaire almost by half, therefore making it less costly to use for the evaluator, while still targeting the most important

aspects of the well-being of the participants. Additionally, it also might increase the validity of this aspect of the evaluation as a shorter questionnaire could increase response rates from participants. The questionnaire remains not part of the outcomes based on which outcome payments are made.

The qualitative aspect of the evaluation has been transformed completely. Based on the tests, it was deemed that it would be better to focus the qualitative aspect of the evaluation on the processes of the SIB, rather than the processes of the service provider. This is different from the initial criteria that were set for the evaluation tool. However, evaluating the processes, is definitely something that can add significant value to the lessons learned of a SIB. Therefore, potentially improving new SIBs. For this a qualitative questionnaire has been developed that should be used for interviews with all directly involved stakeholders, but at least one representative of each investor, commissioner and service provider.

Monitoring has been set at once every three months and now fully focuses on the outputs and outcomes of the SIB. For the same reasons as in the evaluation part, the qualitative evaluation of the service provider has been removed from the tool. There is no point of evaluating the SIB processes in an ongoing manner though, as this is most likely a relatively static process. Therefore, the monitoring tool should collect hard data on the outputs and outcomes of all participants currently active in the SIB and report that to all involved stakeholders. This report is then part of a meeting between all stakeholders in which the progress is discussed and where needed, slight adjustment can be made based on the data.

While the evaluation tool has been designed, tested and redesigned, it should also be reflected on. Working on social issues is complex and different from for instance doing a lab experiment. Consequentially, such a tool can never be perfect, just as SIBs won't ever be perfect. As discussed previously this has its effect on the causality and the validity of the tool. While measures have been taken to make sure that the evaluation is valid and will produce unbiased results, it is unlikely that in all cases exogenous effects can be identified and controlled for. It is questionable whether an evaluation tool for SIBs can ever establish that though and therefore limitations to proving causality of a SIB is something that most likely should be recognised by stakeholders of SIBs.

Another limitation that the tool provides is the evaluation of unintended outcomes. The most important aspect for all stakeholders will be evaluating the direct effects, the intended outcomes. This is what outcome payments are based upon and what is striven for from a social perspective. It is likely though, that social interventions will also have other effects, positive and negative, on the service users which are not always foreseen. The tool tries to catch those effects by looking at the impact the service has on the participant's and identifying the effects that are different from those that were foreseen. But the combination with the potential exogenous effects might limit the possibility to identify all unintended outcomes. It can be possible that not even the participants will know whether a change was because of the provided intervention or something else. Therefore there will obviously be a limit on the unintended outcomes that can be identified.

The relevance of this research for SIBs and potentially all forms of responsible innovation that aim to achieve social impact, remains important though. SIBs are complex, but bring significant opportunities for the field of responsible innovation, but also for governments, investors and social entrepreneurs that are separately trying to do their part in improving the world. These opportunities need to be evaluated correctly to allow them to improve, scale and gain broader recognition. This evaluation tool provides the opportunity for SIBs to be evaluated, analysed and improved.

Overall, this thesis has gone through the design cycle and provided a generic evaluation tool for Social Impact Bonds. While evaluating the outcomes will be specific to the different SIBs, the other aspects of the evaluation tool have been made generic. Therefore, the tool provides transparency and practical achievability. It also allows for comparison of the impact and the processes of the SIBs amongst each other, increasing the learnings that can be taken from SIBs and support further development of the product SIBs in order to reach bigger scale and create more social impact.

8.1. Recommendations

This thesis has provided a first step towards improved evaluation of SIBs. To the knowledge of the author it has been the first study on its evaluation, for which the goal was to design and test an evaluation tool for Social Impact Bonds. The design provides a tool that can not only be used in practice, it would also allow for the comparison of SIBs and therefore be a next step of the development of SIBs as a product.

As can be seen in the introduction of this thesis, SIBs were started as a practical solution to an issue and is not necessarily based on grounded theory. While this does not have to be a bad thing in any circumstance, it also doesn't mean that it can't benefit from more theoretical research. Just as SIBs provide opportunities for the public sector to learn from the private sector and vice versa, SIBs can benefit from different aspects of research. This means that in general there is lots of room for further research. In the upcoming years, the first results of SIBs will start to come through, opening up the system for investigation on its effectiveness. Finding out to what extent SIBs have delivered more or better outcomes than regular subsidized programmes or maybe even fully privately funded programmes, will be crucial to see whether SIBs indeed are a good instrument to deliver more societal impact in the future.

For this specific research there are also various aspects that remain open for further investigation. Several items have not been in the scope of this thesis, but would be considered as important follow-ups of this study. While this thesis has looked at the design of a generic evaluation tool, it has not been able to test this in practice on a larger scope. This design should be taken and used on active SIBs, looking at the practical achievability of the evaluation tool, but also looking at the comparability of different SIBs. This would require a larger study in which several SIBs can go through a full evaluation process, ideally starting directly at the beginning of the SIBs so that also the monitoring and baseline collection can be executed correctly. Furthermore, SIBs that have already ran should be investigated on the adaptive decision making and the extent to what that has been used in existing SIBs. Also there is significant room for the investigation of the governance model of a SIB, what effects have the different incentives had on the acting of stakeholders within a SIB, has it affected outcomes and how?

New research opportunities appear to be endless for such a new initiative that hasn't had large scale studies performed thus far. If SIBs continue to scale as they have been doing in the past years, it is a matter of time before those studies will be carried out as well. It is then up to both the researchers and the marketplace for SIBs to make best use of each other and continue to learn and improve.

Evaluation tool

The evaluation tool consists of two parts, evaluation and monitoring. It will also discuss the who and when question in the governance section.

Governance

The evaluation should be performed by a third party organisation that is not directly involved in the SIB. The fee for the evaluator should not be reliant on the success of the SIB and should be covered evenly by all direct stakeholders of the SIB. The evaluator should have proven expertise in evaluating projects as well as in working in complex social environments.

All direct stakeholders should engage with the outcomes and the proceedings of the evaluation. Quarterly for monitoring meetings in which the results are discussed, agreed upon (agreement should be unanimously) and decision are made around potential changes based on the learnings provided by the monitoring events. The final discussion of the evaluation aspect should take place after every cohort of participants, in the case of multiple cohorts, and at the end of the SIB.

During the SIB, service providers and commissioners should support the evaluation by ensuring that the evaluator has access to all the information/data it needs in order to properly perform the evaluation. In the case of the service provider, this also involves middling towards participants to ensure their participation in the evaluation of their well-being.

Evaluation

The evaluation aspect consists of three aspects. Evaluating the outcomes of the SIB itself, assessing the impact it has had on the SIB participants and qualitative evaluation of the SIB's processes. Their contents are discussed separately. The aim of the evaluation is two sided. The main goal is to identify the effects of the SIB on its participants and its social impact. This is mainly done by evaluating the outcomes and the impact on participants. The qualitative evaluation has a broader goal to provide deeper insights into the processes of the SIB and how they can be improved in future occasions.

Evaluating outcomes

As said, this part will remain practically the same to the initial design. That design recognised that outcomes will be different for all SIBs, as outcome indicators will be set in alignment with the goals of the individual SIB. Nonetheless, the criteria for those indicators can be set beforehand, to ensure that appropriate indicators are being used and evaluation can take place properly. It is restated by the interviewed stakeholders that it is important that the indicators that are being used are agreed upon and objective, apolitical. This will be therefore be maintained in the design. Evaluating outcomes should follow the same four steps:

1. *Determine criteria for baseline data* While a sizeable group of SIBs have been using a control group, based on the conclusions of the previous chapter it is still advised to use baseline data for the evaluation of the outcomes. While the criteria that are used to collect the baseline data should make sure that the baseline represents the participants as much as possible, it is also clear that due to regulations

or sometimes just general poor data collection, it won't always be possible to do exactly that. This should either be assumed through a negotiation process as initially suggested, but potentially the most accurate solution would be to let the evaluator do research on those factors and propose a controlling method for those based on existing data and/or literature.

2. *Collect data at start of the intervention for baseline and participating service users* Nothing has changed to this step of the evaluation. First the starting position of the participants should be collected, but then the baseline should also be collected for that specific group of participants based on their characteristics. The baseline should be as current as possible, but it is recognised that sometimes the baseline will be set based on historical data.

3. *Collect data after the intervention has completed for baseline and participating service users* The same process as in step 2 will take place but then directly after the intervention. When the baseline data can also be reassessed this should be done to be able to control better for exogenous effects.

4. *Analyse data to determine final outcomes* When baseline data could be collected before and after the intervention, the analysis will be different than when baseline data could be collected only once. In the first instance, the achieved outcomes should be compared with the baseline and subsequently corrected for the change in the baseline to correct for the exogenous effects that have taken place during the intervention. If not, the outcomes of the participants should simply be compared to the baseline.

The evaluation of the outcomes will purely look at the intended outcomes for which outcomes are constructed, although the same approach can also be used for broader foreseen outcomes. When possible it will also look at the exogenous effects of the SIB system though, but this will mostly be possible for indicators allowed for the collection of baseline data before and after the intervention. It can obviously be the case that baseline data can be collected before and after the intervention for certain indicators, but not for others. In those circumstances the correction for the exogenous effects should only take place for those indicators where they can be established.

Evaluating impact on participants

Based on the Capabilities approach and the analysis of existing SIBs a questionnaire has been constructed. This questionnaire should be completed by each participant of the SIB before the start of the programme and after the programme is completed.

Introduction: Thank you for agreeing to take part in this survey. We are trying to get a better understanding how the programme of [Service Provider] has effect on you as a person and we will use this survey to gain your thoughts and insights on your perspective on certain aspects of your life. All answers will be collected anonymously and will be kept in the strictest confidentiality. In total the survey should take approximately 10 minutes to complete.

General questions:

- What is your month and year of birth?
 - Month:
 - Year:
- What is your gender?
 - Male
 - Female
 - Other (i.e. Transgender)
 - Prefer not to say
- What is your current country of residence?

Questions related to the person's capabilities:

- Do you eat meat, chicken or fish at least twice a week?
 - Yes
 - No

- If you could choose, would you stay here in your present home or would you prefer to move somewhere else?
 - Yes
 - No
- Does your health in any way limit your daily activities compared to most people of your age?
 - Yes
 - No
- Does your health keep you from doing some types of work?
 - Yes
 - No
 - Prefer not to Say
- In general terms, how would you describe your health?
 1. Not healthy at all
 2. Slightly unhealthy
 3. Adequately healthy
 4. Above average healthy
 5. Very healthy
- Do you feel that on any given day, you could go wherever you want?
 - Yes
 - No, because I cannot afford it
 - No, for other reasons
- Does your accommodation have vandalism or crime in the area?
 - Yes
 - No
- Have you ever been the victim of violent or sexual assault, domestic violence, or any other form of attack?
 - Yes
 - No
 - Prefer not to say
- How likely do you think it is that you will be a victim of violent or sexual assault, domestic violence, or any other form of attack in the future?
 - 1 - Very unlikely, to 7 - Very likely
- How likely do you think it is that within the next 12 months you will be stopped and searched by the police when it is not warranted
 - 1 - Very unlikely, to 7 - very likely
- What is the highest educational qualification that you have achieved until now?
 - None
 - Elementary school
 - Secondary school diploma (including GCSE results)
 - Sixth Form (including A-levels)
 - Further education other than University
 - Bachelor's degree from a university

- Post-graduate qualifications
 - Prefer not to say
- What is the highest educational qualification that, if you had sufficient time and money, you think you could be able to achieve?
 - None
 - Elementary school
 - Secondary school diploma (including GCSE results)
 - Sixth Form (including A-levels)
 - Further education other than University
 - Bachelor's degree from a university
 - Post-graduate qualifications
 - Prefer not to say
- How likely do you think it is that you will achieve that educational qualification?
 - 1 - Very unlikely, to 7 - Very likely
- Do you feel free to express your political views?
 - Yes
 - No
 - Prefer not to Say
- Do you feel free to practice your religion the way you would want to?
 - Yes
 - No
 - Prefer not to Say
- Do you and/or your household have friends or family for a drink or meal at least once a month?
 - Yes
 - No
 - Prefer not to Say
- Would you like to be able to have friends or family for a drink or meal at least once a month, but must do without because you cannot afford it?
 - Yes
 - No
 - Prefer not to Say
- Have you recently been thinking of yourself as a worthless person?
 1. Not at all
 2. Less than usual
 3. Slightly more than usual
 4. Much more than usual
- To what extent do you feel that you respect, value and appreciate other people, independent of their race, sex or religion?
 - 1 - Not at all, to 7 - I fully respect other people
- Are you currently employed?
 - Yes
 - No
 - Prefer not to Say
- Have you ever lost a job at any time and remained unemployed for at least 6 consecutive months? (not when you were a first-job seeker)

- Yes
 - No
 - Prefer not to Say
- If you are currently unemployed, would you like to find a job in the next twelve months?
 - Yes
 - No
 - Prefer not to Say
- If you are currently employed, would you like to switch jobs in the twelve months?
 - Yes
 - No
 - Prefer not to Say
- How likely do you think it is that, if you currently were to look for a (new) job, that you would be able to find a job?
 - 1 - Very unlikely, to 7 - Very likely
- When seeking work in the future, how likely do you think it is that you will experience discrimination because of your:
 - Race (1 - Very unlikely, to 7 - Very likely)
 - Sexual orientation (1 - Very unlikely, to 7 - Very likely)
 - Gender (1 - Very unlikely, to 7 - Very likely)
 - Religion (1 - Very unlikely, to 7 - Very likely)
 - Age (1 - Very unlikely, to 7 - Very likely)
- Outside of any work or employment situation, how likely do you think it is that in the future you will experience discrimination because of your:
 - Race (1 - Very unlikely, to 7 - Very likely)
 - Sexual orientation (1 - Very unlikely, to 7 - Very likely)
 - Gender (1 - Very unlikely, to 7 - Very likely)
 - Religion (1 - Very unlikely, to 7 - Very likely)
 - Age (1 - Very unlikely, to 7 - Very likely)
- Do you have parents or children whom you see less than once a month?
 - Yes
 - No
 - Prefer not to Say
- Are you currently caring for a person who is residing with you (children, parents, or other)?
 - Yes
 - No
 - Prefer not to Say
- Are there any people who you would want to be caring for, but cannot care for them for any reason?
 - Yes
 - No
 - Prefer not to Say
- Are you currently residing with someone who is caring for you?
 - Yes
 - No
 - Prefer not to Say
- Are there any people that you would like to have caring for you, but you are not residing with for any reason?

- Yes
- No
- Prefer not to Say

- Have you recently felt capable of making decisions about crucial parts of your life?
 1. Much less capable than usual
 2. Less capable than usual
 3. More capable than usual
 4. Much more capable than usual
- I have a clear plan of how I would like life to be?
 - 1 - Not at all, to 7 - I know very clearly how I would like life to be
- Did you vote in the most recent general election?
 - Yes
 - I was not able to vote due to my age
 - I was not able to vote due to other reasons
 - I did not vote
- For which of the following reasons, if any, have you not bought your home?
 - Forced, not for particular reasons
 - Affordability
 - Difficulty obtaining mortgage
 - Chose not to buy for other reasons
 - I am a home owner

Closing: Many thanks for completing the questionnaire. Please use a provided envelope to include your answers in or click on finish, in case you answered the electronic questionnaire. Your time and answers are much appreciated and will be processed anonymously and with the strictest confidentiality.

Qualitative evaluation

The questionnaire of the qualitative evaluation will be used after the completion of the SIB and is taken with all direct stakeholders of the SIB. At least one representative of each investor, commissioner and service provider should complete the questionnaire.

1. Why did you participate in this Social Impact Bond?
2. As an organisation, have you benefited from the SIB?
3. Do you think the overall goals of the SIB have been realised?
 - If not, why?
4. Do you think your individual goals for this SIB have been realised?
 - If not, why?
5. How could you have benefited more from the SIB?
6. How have the relationships with the other stakeholders developed?
7. How could the ... have benefited more from the SIB?
 - Commissioner
 - Investor
 - Service Provider

- Participant
8. How would you have made the agreements with the other stakeholders differently than set out in this SIB, using what you have learned in this SIB?
 9. What outcomes were the result of this SIB that you hadn't foreseen at the start?
 10. In hindsight, what changes would you have made to the indicators that have been used for this SIB?
 11. Overall, could you explain which parts of the SIB could be improved and why?

Monitoring

Monitoring should take place quarterly and the results of the monitoring exercise should be discussed with in a monitoring meeting with all direct stakeholders present. While monitoring is something that can give the stakeholders an accurate view of the progress of the SIB, there is also an important aspect to the monitoring meetings. In those meetings, the progress is discussed but the primary goal of the monitoring aspect is to allow for adaptive decision making during the SIB in cases where the monitoring outcomes suggest that changes can be made to improve the tool. Additional information might be agreed upon for the monitoring, but at least the following items should be reported and discussed every three months:

- All participants that are currently participating in the intervention
- When relevant: the stage current participants are in and what part of the intervention they are involved in
- A comparison of the current number of participants and the expected number of participants until that point in time
- Projection of expected outcomes based on the number of participants, the stage they are in and the expected success rates
- A comparison of the total expected outcomes as currently standing and the negotiated expected outcomes as set when the SIB started
- A financial overview of the expected return, based on the achieved outcomes and the projection of expected outcomes

B

Interview responses

All responses in interviews related to evaluation, codified and summed

Response (type)	# Mentioned
Monitoring is an important aspect of evaluation that should take place frequently	6
The most important aspect of evaluation is to look at the outcomes of the SIB	5
Evaluating well-being of participants would add value to the evaluation	5
Valuation of the outcomes is primarily the job of the commissioner	4
Investors should have a certain level of trust in the service providers	4
Adjusting agreements during the SIB is not possible, the agreements should stand	3
There are no incentives to evaluate the processes of the service provider	3
Evaluation should look at the actual impact a programme has	3
Softer outcomes are important for the impact of the SIB, but relatively hard to measure	3
The experiences of the stakeholders should be evaluated afterwards, look at the processes of the SIB itself	3
Are the outcome indicators even measuring the right things	3
Monitoring once per month is too frequently	2
Valuation of soft outcomes is too dependent on the political environment	2
It is important that the evaluation is independent	2
A third party is needed to conduct an evaluation	2
Evaluating the budgets of the SIB doesn't add value	2
SIBs should not just look at the governmental savings, but also at the broader societal value that is being created	2
A performance manager is needed to ensure active management and monitoring of the programme	2
Unintended outcomes are important to take into account	2
It is important to evaluate whether the SIB model was correct	2
Regulations might cause limitations to the availability of data	2
Most exogenous effects cannot be taken into account at all	2
Monitoring in detail would also make the SIB more expensive	2
Monitoring should look at the outcome indicators	2
The processes of the service provider cannot be judged by the other stakeholders due to a lack of expertise	2
Each stakeholder, including evaluator, should focus on what it is good at in the process of the SIB and not on other aspects	2
Sometimes more in-depth analysis is not being done until the results are worse than expected	2
A SIB is an expensive product, so it is important to limit additional costs	2
It is unclear who would have an incentive to evaluate changes in the well-being of participants	2
It is important that the right participants are selected	2
Allowing for a change in the agreements midway, would increase the risks for investors	1
Hard indicators should be leading in constructing a business case for a SIB	1

Response (type)	# Mentioned
Qualitative evaluation would be a good addition to current evaluations	1
A SIB also provides the opportunity for stakeholders to learn from each other	1
Sometimes there are a lot of legal agreements within the contracts of a SIB, but not all of those agreements are being used when the opportunity is there	1
An evaluator should work closely with the SIB	1
An evaluator should not work too closely with the SIB	1
The SIBs should consider drop-outs better in determining success rates	1
When possible, exogenous effects should be taken into account in the evaluation	1
Qualitative factors should not be leading in an evaluation	1
Feedback of the participants on the intervention is important, but primarily for the service provider, not directly for the SIB	1
There can't be too much time between the intervention and its effects	1
Data analysis is critical in the evaluation process	1
An evaluation tool cannot be fully generic, due to the differences between the SIBs	1
A control group is interesting, but would also provide disadvantages	1
Experts should test the model, and its evaluation beforehand on the criteria	1
Outcomes should not just look at the short term, but also at the long term effects	1
People tend to be more honest after the programme has been completed than before	1
It is important that the evaluation itself doesn't create incentives that could lead towards biases	1

Bibliography

- [1] 3SC Capitalise. 3sc capitalise - welsh social impact bond in action. <http://www.3sc.org/news/3sc-capitalise-welsh-social-impact-bond-action>, June 2013. Extracted on: 05/09/2015.
- [2] Sarah Adam. Evaluating social fund impact: A toolkit for task teams and social fund managers. *SP Discussion Paper*, (0611), October 2006.
- [3] Sabina Alkire. Why the capability approach? 1998.
- [4] Sabina Alkire. Choosing dimensions: the capability approach and multidimensional poverty. *Chronic Poverty Research Centre Working Paper*, (88), August 2007.
- [5] Paul Anand and Martin van Hees. Capabilities and achievements: An empirical study. *The Journal of Socio-Economics*, 35:268–284, 2006.
- [6] Paul Anand, Graham Hunter, and Ron Smith. Capabilities and wellbeing: Evidence based on the sen-nussbaum approach to welfare. *Open Discussion Papers in Economics*, (47), January 2004.
- [7] Paul Anand, Christina Santos, and Ron Smith. The measurement of capabilities. *Open Discussion Papers in Economics*, November 2007.
- [8] James Anderson and Andrea Phillips. What we learned from the nation’s first social impact bond. http://www.huffingtonpost.com/james-anderson/what-we-learned-from-the-_1_b_7710272.html, February 2015. Extracted on: 28/09/2015.
- [9] Christian Arnsperger and Yanis Varoufakis. What is neoclassical economics? the three axioms responsible for its theoretical oeuvre, practical irrelevance and, thus, discursive. *Panoeconomicus*, 1:5–18, 2006.
- [10] A.B. Atkinson. Measuring poverty and differences in family composition. *Economica*, 59(233): 1–16, February 1992.
- [11] David G. Blanchflower and Andrew J. Oswald. Well-being over time in britain and the usa. *Journal of Public Economics*, 88:1359–1386, 2004.
- [12] Andrea Brandolini and Giovanni D’Alessio. Measuring well-being in the functioning space. October 1998.
- [13] Bridges Ventures. Fusing housing uses a multi-disciplinary team approach to support and guidance to reduce youth homelessness. <http://www.bridgesventures.com/portfoliolist/fusion-housing-fusion-fair-chance-partnership/>. Extracted on: 22/09/2015.
- [14] Bridges Ventures. Bridges-backed career connect delivers key sib milestone. <http://www.bridgesventures.com/bridges-backed-career-connect-delivers-key-sib-milestone/>, July 2015. Extracted on: 01/09/2015.
- [15] Richard Brooks. *Out of Sight - How we lost track of thousands of NEETs and how we can transform their prospects*. Fabian Society, London, The United Kingdom, 2014.
- [16] Cabinet Office. Department for work and pensions innovation fund. http://data.gov.uk/sib_knowledge_box/department-work-and-pensions-innovation-fund, April 2013. Extracted on: 06/09/2015.
- [17] Cabinet Office. Fair chance fund - full bid specification documentation. June 2014.

- [18] City of Chicago. Mayor emanuel announces expansion of pre-k to more than 2,600 chicao public school children. <http://www.payforsuccess.org/sites/default/files/chicago-sib-fact-sheet.pdf>, October 2014. Extracted on: 06/09/2015.
- [19] Andrew E. Clark and Andrew J. Oswald. Unhappiness and unemployment. *The Economic Journal*, 104(424):648–659, May 1994.
- [20] David A. Clark. The capability approach: Its development, critiques and recent advances. *ESRC Global Poverty Research Group*.
- [21] Flavio Comim. Operationalizing sen’s capability approach. June 2001.
- [22] European Commission. Responsible research & innovation. <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation>, n.d. Extracted on: 31-01-2016.
- [23] Joppe de Ree and Rob Alessie. Life satisfaction and age: Dealing with underidentification in age-period-cohort models. *Social Science & Medicine*, 73(1):177–182, July 2011.
- [24] Tjalling de Vries. Interview, December 2015.
- [25] Department for Work & Pensions. Innovation fund pilots qualitative evaluation - early implementation findings. *Government Social Research*, (880), July 2014.
- [26] Department for Work and Pensions and The Rt Hon Iain Duncan Smith. G8 social impact investment conference. <https://www.gov.uk/government/speeches/g8-social-impact-investment-conference>, June 2013. Extracted on: 01/04/2015.
- [27] Oxford Dictionaries. British & world english – evaluation. <http://www.oxforddictionaries.com/definition/english/evaluation>. Extracted on: 08/11/2015.
- [28] Bert Enserink. Interview, November 2015.
- [29] Enterprise Community. Fact sheet: the cuyahoga partnering for family success program. December 2014.
- [30] Forbes. The world’s billionaires. <http://www.forbes.com/billionaires/list/>, 2015. Extracted on: 08/07/2015.
- [31] Megan Golden and Brian Nagendra. Pay for success in the u.s. February 2015.
- [32] Massachusetts Government. Massachusetts launches pay for success initiative to reduce chronic individual homelessness. December 2014.
- [33] Emily Gustafsson-Wright, Sophie Gardiner, and Vidya Putcha. *The Potential and Limitations of Impact Bonds - Lessons from the first five years of experience worldwide*. The Brookings Institution, Washington, 2015.
- [34] D. Hardoon. Wealth: Having it all and wanting more. *Oxfam Novib*, 2015.
- [35] Emma Hollywood, Valerie Egdell, Ronald McQuaid, and Dirk Michel-Schertges. Methodological issues in operationalising the capability approach in empirical research: an example of cross-country research on youth unemployment in the eu. *Social Work & Society*, 10(1), 2012.
- [36] Paul Hopkins. Interview, September 2015.
- [37] Christopher Houghton Budd. *Finance at the Threshold - Rethinking the Real and Financial Economies*. Gower Publishing Limited, 2011.
- [38] Key Fund. Social impact report. 2015.
- [39] Koekoek2015. Interview, December 2015.

- [40] Jon Krosnick and Stanley Presser. *Question and Questionnaire Design*, chapter 9, pages 263–314. Emerald Group Publishing Limited, second edition, 2010.
- [41] Stefan Kuhlmann. Evaluation of research and innovation policies: a discussion of trends with examples from germany. *International Journal of Technology Management*, 26(2/3/4):131–149, 2003.
- [42] Enrica Chiappero Martinetti. A multidimensional assessment of well-being based on sen’s functioning approach. *Rivista Internazionale di Scienze Sociali*, (2).
- [43] Geoff Mulgan. Measuring social value. http://ssir.org/articles/entry/measuring_social_value, 2010. Extracted on: 17/11/2015.
- [44] Cecilia Muñoz and Robert Gordon. Pay for success: A new results-oriented federal commitment for underserved americans. <https://www.whitehouse.gov/blog/2012/01/24/pay-success-new-results-oriented-federal-commitment-underserved-americans>, January 2012.
- [45] New York State. Investing in what works: ‘pay for success’ in new york state. March 2014.
- [46] Nottingham Futures. Ground-breaking project in nottingham leads to significant reduction in youth unemployment. <http://www.futuresadvice.co.uk/news/2015/06/18/ground-breaking-project-in-nottingham-leads-to-significant-reduction-in-youth-unemployment/>, June 2015. Extracted on: 03/09/2015.
- [47] Martha C. Nussbaum. *Women and Human Development*. Cambridge University Press, 2000.
- [48] Martha C. Nussbaum. Capabilities as fundamental entitlements: Sen and social justice. *Feminist Economics*, 9(2/3):33–59, 2003.
- [49] Martha C. Nussbaum. *Creating Capabilities - The Human Development Approach*. Belknap press, 2011.
- [50] OECD. *OECD Guidelines on Measuring Subjective Well-being*. OECD Publishing, 2013.
- [51] P3. £1.45million boost for accommodation, education and employment in gloucestershire. <http://www.p3charity.org/News/145million-boost-for-accommodation-education-and-employment-in-gloucestershire>. Extracted on: 22/09/2015.
- [52] Dom Pates. Three faces of london - st. paul’s, millenium bridge and a homeless man. <http://www.patesonline.net/>, March 2008. Used under CC BY.
- [53] Rita Perakis. First development impact bond is launched. <http://www.cgdev.org/blog/first-development-impact-bond-launched>, June 2014. Extracted on: 06/09/2015.
- [54] Udo Pesch. Interview, November 2015.
- [55] The Giving Pledge. The giving pledge. <http://givingpledge.org/>.
- [56] Ian Popham. Tackling neets – research on actions and other factors that can contribute to a reduction in the numbers of young people not in education, employment [neet]. 2003.
- [57] Janette Powell. Interview, October 2015.
- [58] Mozaffar Qizilbash. Poverty: Concept and measurement. *Research Report Series*, (12), 1998.
- [59] Ingrid Robeyns. Sen’s capability approach and gender inequality: Selecting relevant capabilities. *Feminist Economics*, 9(2-3):61–92, 2003.
- [60] Ingrid Robeyns. The capability approach. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2011 edition, 2011.
- [61] Ruhi Saith. Capabilities: the concept and its operationalisation. *QEH Working Paper Series*, 2001.

- [62] Yasemin Saltuk, Ali El Idrissi, Amit Bouri, Abhilash Mudaliar, and Hannah Schiff. Eyes on the horizon – the impact investor survey. May 2015.
- [63] Erik Schokkaert and Luc van Ootegem. Sen’s concept of the living standard applied to the belgian unemployed. *Louvain Economic Review*, 56(3/4):429–450, 1990.
- [64] Amartya Sen. *Commodities and Capabilities*. Elsevier Science, Amsterdam, Netherlands, 1985.
- [65] Amartya Sen. Capability and well-being. In Martha Nussbaum and Amartya Sen, editors, *The quality of Life*. Oxford University Press, 1993.
- [66] Amartya Sen. *Development as Freedom*. Knopf, New York, 1999.
- [67] Amartya Sen. Capabilities, lists, and public reason: continuing the conversation. *Feminist economics*, 10(3):77–80, 2004.
- [68] Amartya Sen. Capability and well-being. In Daniel M. Hausman, editor, *The philosophy of economics*, chapter 15, pages 270–294. Cambridge University Press, New York, 3 edition, 2007.
- [69] Sonal Shah and Kristina Costa. Social impact bonds: White house budget drives pay for success and social impact bonds forward. <https://www.americanprogress.org/issues/economy/news/2013/04/23/61163/white-house-budget-drives-pay-for-success-and-social-impact-bonds-forward/>, April 2013.
- [70] Social Finance UK. Sib database. <http://www.socialfinance.org.uk/ngo-map/>. Extracted on: 03/09/2015.
- [71] Social Finance UK. Social impact bonds - rethinking finance for social outcomes. August 2009.
- [72] Social Finance UK. The essex social impact bond. a year in review. November 2014.
- [73] Social Finance UK. Peterborough social impact bond reduces reoffending by 8.4%; investors on course for payment in 2016. <http://www.socialfinance.org.uk/peterborough-social-impact-bond-reduces-reoffending-by-8-4-investors-on-course-for-payment-in-2016/>. August 2014. Extracted on: 28/09/2015.
- [74] Social Finance US. Social finance drives landmark new york state deal. <http://socialfinanceus.org/what-we-do/select-current-engagements/social-finance-drives-landmark-new-york-state-deal>. Extracted on: 28/09/2015.
- [75] Social Ventures. Case study: Sweet dreams. December 2014.
- [76] Social Ventures Australia. Newpin social benefit bond amendments. July 2015.
- [77] St. Basils. Rewriting futures – helping homeless young people with complex needs. April 2015.
- [78] Mark S. Stein. Nussbaum: A utilitarian critique. *Boston College Law Review*, 50:489–531, 2009.
- [79] Adam Stewart. The perth ymca sib and the role of involved investors. <http://financeforgood.ca/perth-ymca-sib-role-involved-investors/>. Extracted on: 05/09/2015.
- [80] Joseph Stiglitz. The price of inequality: How today’s divided society engangers our future. *W.W. Norton & Company*, 2012.
- [81] The Commonwealth of Massachusetts and Massachusetts alliance for supportive housing LLC. Pay for success contract. December 2014.
- [82] The Y. Ambition east midlands. <http://www.leicesterymca.co.uk/housing-ambition.php>, 2015. Extracted on: 22/09/2015.
- [83] Emma Tomkinson. Social impact bond (sib) in augsburg, germany. <http://emmatomkinson.com/2014/08/20/social-impact-bond-sib-in-augsburg-germany/>, August 2014. Extracted on: 02/09/2015.

- [84] Roger Tourangeau and Ting Yan. Sensitive questions in surveys. *Psychological Bulletin*, 133(5): 859–883, October 2007.
- [85] Kat-Yuen Tsui. Multidimensional generalizations of the relative and absolute inequality indices: The atkinson-kolm-sen approach. *Journal of Economic Theory*, 67:251–265, 1995.
- [86] UBS Optimus Foundation. Investing for social outcomes; children’s investment fund foundation and ubs optimus foundation launch first development impact bond in education. June 2014.
- [87] United Nations Development Programme. Human development report 2014 - sustaining human process: Reducing vulnerabilities and building resilience. 2014.
- [88] United Way of Salt Lake. United way of salt lake announces results-based financing for low-income preschool students. June 2013.
- [89] Christian van Esch. Interview, December 2015.
- [90] Bart van Hulst. Interview, November 2015.
- [91] Leo van Loon. Interview, December 2015.
- [92] Ruut Veenhoven. Advances in understanding happiness. *Revue Québécoise de Psychologie*, 18: 29–74, 1997.
- [93] Björn Vennema and Ruben Koekoek. Social impact bonds - introductie van een nieuw financieringsmodel in nederland, December 2013.
- [94] Joan Miquel Verd and Marti Lopez. The rewards of a qualitative approach to life-course research. the example of the effects of social protection policies on career paths. *Forum: Qualitative Social Research*, 12(3), September 2011.
- [95] Polly Vizard. Developing and agreeing a capability list in the british context: What can be learnt from social survey data on ‘rights’? November 2010.
- [96] Polly Vizard and Tania Burchardt. Developing a capability list: Final recommendations of the equalities review steering group on measurement. April 2007.
- [97] VONNE. Ways to wellness frequently asked questions. <http://www.vonne.org.uk/ways-wellness-frequently-asked-questions>, January 2014. Extracted on: 28/09/2015.
- [98] Ellie Ward. Triple triumph: three uk sibs return investor capital. <http://www.pioneerspost.com/news-views/20150715/triple-threat-three-uk-social-impact-bonds-return-investor-capital>, July 2015. Extracted on: 02/09/2015.
- [99] Gianni Zappalà and Mark Lyons. Recent approaches to measuring social impact in the third sector: An overview. *CSI Background Paper*, (6), July 2009.