

Measuring well-being and sustainability: irreconcilably opposed or inevitably linked?



Utrecht University

Abstract: *This thesis is an inquiry into the complexity of well-being measurement, especially when attempting to incorporate sustainability. To analyze the dynamics between these two contested concepts in social sciences, I compare three measures of (sustainable) well-being: the Brede Welvaartsindicator, the Sustainable Development Goals and the Monitor Brede Welvaart. The three steps of measurement developed by Cartwright and Runhardt (2014) guide the analysis, which is performed through desk research, semi-structured interviews and a newly developed typology of well-being indicators. The analysis shows that each of the current measures of (sustainable) well-being is imperfect. Using this, I identify four fundamental problems which any researcher in well-being measurement must relate to. Finally, I translate the lessons learned into practical recommendations for the improvement of the Brede Welvaartsindicator. Most importantly, I argue that this measure would benefit from the inclusion of capital stock indicators to represent the sustainability of current well-being in relation to future generations.*

Keywords: well-being, sustainability, measurement, economic methodology, case-study analysis

Final thesis U.S.E.
MSc Economic Policy
June 29, 2022
Wordcount: 15657

Harold Janssen (0946494)
h.j.janssen@students.uu.nl
Supervisor: Prof. dr. ir. Marcel Boumans
m.j.boumans@uu.nl
Second reader: Dr. Loek Groot
L.F.M.Groot@uu.nl

Table of Contents

Preface	3
Introduction	4
Literature review	6
Measurement in social sciences	7
A history of definitions: well-being, sustainability and sustainable well-being	7
A history of measurement: replicability and presentation	9
Theoretical framework	11
Three steps of measurement: characterization, representation and procedures	11
Challenges to the three steps: dashboards and index-making	12
Challenges to the three steps: defining and discounting (future) well-being	14
Methodology	16
Case study selection: the BWI, the SDGs and the MBW	17
Case-study analysis: desk research and the typology of indicators	19
Case-study analysis: expert interviews	20
Analysis - Individual cases: characterization, representation, procedure	21
Brede Welvaartsindicator	22
Sustainable Development Goals	24
Monitor Brede Welvaart	27
Analysis – Comparative analysis: Real-world effects and fundamental problems	29
Making an impact	30
Four fundamental problems of well-being and sustainability measurement	31
Problem 1. The problem of defining a living concept	32
Problem 2. The problem of accuracy	32
Problem 3. The problem of future preferences and capabilities	33
Problem 4. The problem of presentation	34
Analysis - Improvements to the BWI: increasing transparency and incorporating sustainability	35
Recommendation 1	35
Recommendation 2	36
Recommendation 3	37
Recommendation 4	37
Concluding remarks	39
A reflection on this research	40
Appendix A. Semi-structured interview questions (example)	42
Bibliography	43

Preface

This thesis marks the end of two years of personal development in the field of economics, first through the premaster programme and then through the master programme in Economic Policy at Utrecht University. The topic of this thesis, measurement issues in well-being and sustainability, is a timely representative of these years. Both the topic and the research have surprised me multiple times along the way.

At first, the issue of how to incorporate sustainability into well-being measurement seemed like a rather small focus for a thesis, since there are only a few examples of such measures available. However, the literature on the topic has been developing rapidly in the past decades. The further I dove into the topic, the more I felt like I had barely scratched the surface of the knowledge already out there.

The case selection allowed me to focus my research on three measures of well-being which were simultaneously quite similar and very different. I enjoyed the in-depth analysis of many publications and felt challenged to find larger patterns which were useful for the thesis. Both the fundamental problems of well-being measurement and the typology of indicators which I propose in this research are an attempt at identifying these larger patterns.

Issues of measurement have existed for as long as economics has been around. This research is just a small contribution to the existing literature on these issues. If anything, I have learned that measurement in social sciences cannot be performed with the objectivity which social scientists might aspire to. However, this does not necessarily constrain social scientists. By being transparent about the choices which have been made and by acknowledging some degree of subjectivity, social scientists can still set a certain standard of how to measure living concepts such as well-being and sustainability.

For this research, I am indebted to the people who helped me along the way. First, I want to thank the three interviewees for their valuable insights into the cases: Maaïke Stoel (Dutch ministry of economic affairs), Rixt de Jong (CBS) and prof. Jan-Luiten van Zanden (Utrecht University). Secondly, I want to thank both my supervisor and my second reader for their advice along the way, especially when it comes to out-of-the-box thinking and keeping the research feasible: Prof. dr. ir. Marcel Boumans and dr. Loek Groot. It has been a valuable experience to write this thesis with the help of so many people experienced in the field of economic methodology, sustainability and well-being.

Harold Janssen

Leiden, June 29th, 2022

1. Introduction

Well-being is trending among Dutch economists, policymakers and citizens. Whether it is in a parliamentary debate about the yearly budget (*Aanhangsel Handelingen II, nr. 81, item 2*, 2022), a popular newspaper (Smit, June 12, 2022) or a trade journal for economists (van Bree and de Jonge, 2022), the importance of people's well-being has grown in the last decade. This development reflects a shift in global economic discourse, away from economic production and towards a more comprehensive view of a country's economic performance. Simultaneously, sustainability as a topic in policy-making has been gaining momentum for decades and has become a contested subject in many national and international political debates. Kordestani, Peighambari and Foster (2015) for example show how between the decades 1991-2000 and 2001-2010, academic publications on sustainability have increased fifteenfold.

While sustainability and well-being have both become more salient, they do not necessarily complement each other. On the contrary, an increase in well-being of people living today can affect well-being of people yet to be born. Depletion of ecological, social and economic capital currently in existence can become a constraint for future generations. Alternatively, an increase in individual well-being in one country can decrease well-being in another. How, if at all, does one measure such intratemporal and intertemporal well-being dynamics? Many different measures exist for well-being, for sustainability and for sustainable well-being. How then do these measures define such living concepts and how do they incorporate well-being trade-offs between now, later and elsewhere?

In this thesis, I analyse how researchers measure the broad concepts of well-being and sustainable well-being. I determine which choices researchers have to make and why they choose one alternative over the other. The main research question of this thesis therefore is: *To what extent can measurement of well-being incorporate sustainability?* Within the scope of this research, it is not possible to analyse the myriads of measures of (sustainable) well-being, welfare and economic production. Instead, a case-study based analysis of three different measures of (sustainable) well-being guides this research. The *Brede Welvaartsindicator* (hereafter: BWI), developed by researchers of Utrecht University and RaboResearch is the main case of interest. It is a measure for Dutch well-being which specifically does not incorporate sustainability (Rijpma et al., 2017). I compare the BWI with two measures of well-being which do incorporate sustainability: The United Nations Sustainable Development Goals (hereafter: SDGs) and the *Monitor Brede Welvaart* (hereafter: MBW). The analysis consists of desktop research, expert interviews and the development of a typology of well-being indicators.

This analysis takes place within a framework developed by Cartwright and Runhardt (2014), the three steps of measurement: characterization, representation and procedure. The application of this framework to the selected cases separates the main research question into six subquestions which I will answer throughout this thesis.

Subquestions

1. How have well-being, sustainability and sustainable well-being historically been defined?
 2. How have well-being, sustainability and sustainable well-being historically been measured?
 3. How do the BWI, SDGs and MBW characterize (sustainable) well-being?
 4. How representative is the set of indicators chosen to measure the BWI, the SDGs and the MBW?
 5. To what extent does the procedure to measure the indicators fit the representation and characterization of (sustainable) well-being?
 6. How do the lessons learned from the SDGs and MBW apply to the incorporation of sustainability into the *Brede Welvaartsindicator*?
-

This research is relevant to academia as well as to society. Academically, an inquiry into historical and current attempts at measuring sustainability and well-being can contribute to the field of economic methodology. Many studies have proposed different approaches of defining and measuring sustainable well-being (e.g. Stiglitz, Sen and Fitoussi, 2009; Costanza et al., 2016). However, this study is a comparative analysis of three approaches with differing purposes, target audiences and methods. Using the framework of Cartwright and Runhardt (2014), I carefully consider the internal consistency of the three cases. By zooming in on three measures and tracking each case from the definition in its first publication to the measurement procedure in the last, the challenges and fundamental problems of measuring complex concepts in social sciences come to the fore. Simultaneously, the theoretical framework of Cartwright and Runhardt (2014) is put to the test: Is it applicable to the concepts of sustainability and well-being? Finally, and arguably most importantly, I make two new contributions to the field of economic methodology by identifying four fundamental problems of well-being measurement and by developing a typology of well-being measurement indicators. I invite other researchers to put these contributions to the test using a different set of cases.

This research is societally relevant because it proposes practical recommendations to improve a Dutch measure of well-being, the BWI. Specifically, it proposes an answer to the challenge identified by the authors of the BWI, namely on how to incorporate sustainability into the measure (Rijpma et al., 2017). Moreover, the comparative analysis of the societal

impact which each measure makes showcases how impact has been made and how future researchers could make impact. The analysis shows that each measure has impacted the societal or academic debate differently. Moreover, the trade-offs identified in measuring well-being and sustainability show how difficult it is to facilitate an evidence-based discussion about concepts such as well-being and sustainability. The analysis shows that it is highly unlikely for a perfect measure of either concept to exist, let alone one measure which combines both sustainability and well-being. Current measures of (sustainable) well-being are necessarily imperfect. If debates in society about these measures start by acknowledging their flaws and by being transparent about the choices that researchers have made, this could increase the quality of the debate. Considering that trade-offs between well-being and sustainability are likely to remain part of public discourse, knowing the limits of the tools through which we measure these concepts is fundamental.

This thesis consists of five more chapters. In chapter 2, I critically reflect on academic literature relevant to this research. In chapter 3, I propose the theoretical framework through which each case will be studied. In chapter 4, I justify the case selection and elaborate on the methodology of the case-study based research. Chapter 5 contains the analysis, the first part of which contains the case specific analyses. In the second part, through comparative analysis, I identify four fundamental problems which researchers face when measuring well-being. I translate these lessons learned to recommendations for improvement of the BWI. Finally, the conclusions and a critical review of these conclusions are presented in Chapter 6.

2. Literature review

Measurement of well-being and of sustainability are well-established topics in the field of economic methodology. A brief exploration of the main debates and their relation to the research questions is therefore required. By placing this exploration in a historical perspective, the literature review provides answers to the first two research questions as given below. This review discusses three subjects: 1) current approaches to measurement in social sciences; 2) historical developments in the concepts ‘sustainability’, ‘well-being’ and ‘sustainable well-being’ and 3) historical developments in measuring these concepts and presenting the results.

-
1. How have well-being, sustainability and sustainable well-being historically been defined?
 2. How have well-being, sustainability and sustainable well-being historically been measured?
-

Measurement in social sciences

Concepts such as sustainability and well-being are not observable to humans like the sound of a drum or the weight of an apple. These concepts require different approaches to measurement than in many fields of hard science. Social scientists are forced to make choices about the definition and subsequent operationalization of the concepts they intend to measure. These choices are subjective, making "mechanical objectivity" unattainable in social sciences (Porter, 1995, p. 7). Due to this inevitable subjectivity, no measure is perfectly right or perfectly wrong. Instead, According to Porter (1995), multiple valid measures for one concept can coexist.

How then, does one measure in social sciences? Zeller and Carmines (1980) discuss the abstractness of concepts in social sciences and argue that abstract concepts can be measured in many ways. Because a concept in social sciences in itself is not measurable, scholars must motivate, test and evaluate the use of valid and reliable indicators. The validity and reliability of these indicators then hinge on the internal and external association, being the relationship of the indicator with the concept and with other indicators respectively (Zeller and Carmines, 1980). Nowakowska (1977) also explores the relationship between the fuzziness of concepts in social sciences and their measurement procedure, attempting to introduce a formal approach to empirical measurement. Similar to Zeller and Carmines (1980), Cartwright and Runhardt (2014) establish that a "systematic and grounded" way of measuring is key (p. 267). Subsequently, they define three steps of measurement: characterization, representation and procedures. The lack of formal notations makes this framework more malleable and more accessible than the frameworks of Zeller and Carmines (1980) and Nowakowska (1977). This makes it most applicable for a case-study approach to sustainability and well-being. These three steps of measurement, which will be elaborated upon in the theoretical framework, have not been applied to (sustainable) well-being, making it a novel approach.

A history of definitions: well-being, sustainability and sustainable well-being

The first of the three steps of measurement, to 'characterize' or define a concept, is a challenge when it comes to 'sustainability', 'well-being' and 'sustainable well-being'. A variety of contested definitions exists for these concepts. Well-being on the individual level is quite a new term in economics. Previously, economists and policymakers had mostly been interested in the production capacity of a country, for example to gauge the potential war time production (Coyle, 2014). Gross Domestic Product (per capita), an indicator for production capacity, functioned as an arguably poor proxy of economic performance and well-being until the introduction of the Human Development Index (HDI). The HDI centred around the formation

and use of human capabilities, an approach developed by Sen (2000). It combined indicators for life expectancy, education and economic performance into one index representing and ordering national well-being of participating nations (UNDP, 1990).

This new foundation substantially broadened the concept of and discussion about well-being for social scientists and for policymakers. They stepped away from their narrow-minded focus on production. The government of Brunei for example has set goals based on HDI: Brunei's objective to become a top 10 HDI ranking member in 2035 is useful in guiding its economic, fiscal and sectoral policy in the country (Michael, 2018). It is far from alone in using HDI to set goals. The HDI gave rise to other, even broader interpretations of well-being. An influential report by Stiglitz, Sen and Fitoussi (2009) built further upon the capabilities approach to well-being, adding that experienced well-being is subjective and that inequality plays a role in each dimension of well-being. This capabilities approach has been of influence to current measures of well-being, such as the BWI and the MBW.

For sustainability, multiple definitions have existed and exist as well. A narrow definition of sustainability entails the depletion of finite natural resources and the pollution of the natural environment. It originated with biologists and ecologists before being introduced into (ecological) economics (Vos, 2007, p. 335). This narrow perspective is for example used by Hamilton (2016) in an attempt to link sustainability to national accounting. Arguably, the focus on environmental sustainability has been dominant in public discourse since the Limits to Growth Report by the Club of Rome (Meadows et al., 1972). However, when economists became concerned with sustainability, more types of sustainability arose: economic sustainability concerns the extent to which economic growth can continue into the future; population sustainability concerns the sustainable growth of the world population; social sustainability is concerned with (intergenerational) equity and redistribution (Vos, 2007; Stiglitz, Sen and Fitoussi, 2009). Different scholars use different 'thick' and 'thin' versions of sustainability. Being aware of these different definitions is important: as definitions differ, operationalization and measurement procedures are likely to differ as well. For this research, I consider sustainability as the extent to which current generations leave behind enough capital and resources for future generations. This contains ecological, economic and social types of capital, such as natural resources, production capital and human capital.

While the definitions of sustainability and well-being remain contested, a shift in economics has taken place towards sustainable development and sustainable well-being. The Brundtland Commission paved the way for this term in 1987, defining sustainable development as "development that meets the needs of the present without compromising the ability of future

generations to meet their own needs.” (World Commission on Environment and Development, 1987, p. 41). Critically reflecting on this definition and similar definitions, Vucetich and Nelson (2010) argue that what this means depends on how people understand normative terms such as 'development' and 'the needs of the present'. They argue that sustainability and sustainable development are "essentially the relationship between the environment and society", consisting of both physical and ethical aspects (p. 540). Kjell (2011) states that these normative questions about the relation between the environment and society make it difficult for researchers to remain objective in the study of sustainability and well-being. To define contested concepts, it remains valuable to openly discuss individual values underlying these definitions (Kjell, 2011). In the analysis, this research briefly considers the role of subjectivity in developing measures of well-being, especially for the MBW.

The varying definitions of the concepts 'well-being', 'sustainability' and 'sustainable well-being' (or 'sustainable development') are emblematic of the continuing debate among academics. One concept has multiple definitions and two similar definitions can belong to different concepts. Therefore, it is important to always critically study the definition of a concept and the way in which it is being measured. This study contributes to the debate on these different definitions and their measurement procedures by considering the types of societal impact made by each measure.

A history of measurement: replicability and presentation

Having defined a certain concept, operationalizing it and finding the right procedure to measure the concept remains a challenge. For over 250 years, economists have been concerned with the measurement of economic production, development, welfare and individual well-being (Boumans, 2019). Throughout these years, many different methods to measure the size of an economy and the welfare or well-being of its individuals have been developed. It is beyond the scope of this study to discuss the many types of indices, indicators and qualitative tools employed for economic measurement, but two key issues are briefly discussed below: the replicability of the measurement procedure and the presentation of results.

The replicability of a procedure across countries can be complex because it requires statistical agencies of varying sizes and qualities to perform the same measurement procedures for a certain concept. While the United Nation Statistical Division (UNSD, 2015) proposes fundamental principles for official statistics, the replicability of measurement procedures remains an important issue. A case study by Morgan (2009) into the National Income Accounting fieldwork by Phyllis Deane in the 1950s shows how difficult it can be to measure

economic activity across different economies and cultures. Moreover, even if a measure is easily replicable across countries, a country may not be able to afford the expensive methods of data collection required. The SDGs explicitly specify the need to “intensify efforts to strengthen statistical capacities in developing countries” (United Nations 2015, p. 15), showing that this problem still influences the availability of data today.

The second problem which economists and statisticians face in measurement is how to report the results. For (sustainable well-being measurement, two opposing schools of thought currently coexist: the index approach and the dashboard approach. The first proposes to produce one index number which summarizes a variety of well-being and/or sustainability dimensions into one measure. Examples of this school are the Human Development Index and the BWI. An argument in favour of this approach is that the result becomes easily comparable across countries and over time. Ravallion (2012) however argues that these 'mashup indices' have significant downsides: the changing data within the composite index, the poorly justified attribution of weights to individual indicators and other trade-offs remain invisible to those using it (p. 1).

Instead, the second school of measuring well-being produces more insightful results according to Ravallion (2012). This dashboard approach does not make a composite of different well-being dimensions but presents the results for all dimensions in one dashboard. Examples of this approach are the SDGs and the MBW. For both measures, multiple dimensions and indicators have been specified and are reported side-by-side, without weighting, standardizing or aggregating the results. Rijpma et al. (2017) oppose this approach, arguing that a dashboard of indicators falls short as a measurement instrument because the results are difficult to communicate, inaccessible to outsiders and vulnerable to cherry picking. Measures adhering to both schools remain today, with the BWI being an index and the SDGs and MBW being a dashboard.

This brief overview of measurement literature, specifically measurement of well-being, sustainability and sustainable well-being, shows that it is quite a difficult task to measure a concept in social sciences, especially in case of such contested concepts. Still, the summary of developments in the field provide an answer to the first and second research question. Many definitions and operationalizations coexist and either fit or do not fit the purpose which a researcher has in mind. Statisticians, economists and policymakers have been developing new characterizations, new indicators and new ways of presenting data. Still, they must remain critical of what is being measured, how it is being measured and why it is being measured when

it comes to (sustainable) well-being. Next, I will elaborate on a framework which helps researchers to make these choices involved in measurement in social sciences.

3. Theoretical framework

A variety of theoretical debates comes into play when one attempts to measure a contested concept such as (sustainable) well-being. This starts by acknowledging that measurement in social sciences cannot be as objective as measurement in hard sciences. Instead, a framework is required within which the concept can be placed. This chapter introduces the framework by Cartwright and Runhardt (2014) which is applied to the cases. Moreover, two approaches to index-making, the axiomatic and instrumental approach, are briefly discussed. Finally, different approaches to well-being measurement come to the fore, as well as considerations relating to intergenerational well-being measurement and the discount rates associated with it. Altogether, this discussion places the analysis of the BWI, the MBW and the SDGs within a theoretical framework.

Three steps of measurement: Characterization, representation and procedures

Cartwright and Runhardt (2014) have proposed a framework of choices to be made when developing a measurement approach with the three steps of measurement: characterization, representation and procedures. For this research, I borrow their approach to test the validity of the three cases. The first step, characterization, is to define a socially constructed concept in a useful way. Due to the ‘fuzziness’ of a contested concept such as ‘well-being’, a careful delineation of what is understood by the concept is required before a researcher can proceed to measure it (Cartwright and Runhardt, 2014, p. 268). For this research, it means that a definition of (sustainable) well-being needs to be clearly stated for each of the three cases. Moreover, this definition needs to be justified from a larger philosophical perspective of what human well-being constitutes.

Once a concept is characterized, the second step in measurement according to Cartwright and Runhardt (2014) is to find a method of representation: finding one or multiple dichotomous, ordinal or continuous variables which capture the essence of the concept as characterized. For example, a ‘democracy’ can be characterized as a country with free and fair elections. The concept could then be represented by a dichotomous variable measuring whether elections match a predetermined set of criteria for freedom and fairness. Alternatively, two different variables could capture the freedom and fairness of the elections respectively. The broader a concept is defined, the more variables (or indicators) are required to represent it. For

this research, the representation of (sustainable) well-being is a key step in each case: to what extent are the indicators used to represent the concept in line with its definition? In all three cases, the measured concept has been split into multiple dimensions and into even more indicators for each dimension. The decisions which must be made in this process point to fundamental problems in measuring contested concepts. Moreover, for the SDGs and the MBW, a set of dimensions, targets and indicators specifically represents the extent to which well-being is sustainable. Understanding how sustainability is represented in these cases can contribute to a proposal for incorporating sustainability into the *Brede Welvaartsindicator*.

The final step of measurement consists of determining the ‘on-the-ground’ procedures through which to measure the concept, often requiring refinement through a feedback loop with the two earlier stages of measurement (Cartwright and Runhardt, 2014). These procedures consist of the methods used and data sources employed to measure the concept. In the example of democracy, the procedure could be to consult the reports by independent election observants. Obtaining such data often is expensive, requiring researchers to weigh two values: the accuracy of the data and the feasibility of collecting it. Cartwright and Runhardt (2014) stipulate that social scientists often depend on imperfect data collected by others and with other research purposes in mind. It is therefore important to balance a pragmatic view to an external data source with a critical view of the purpose for which it was collected. For this research, it is important to determine how accurately the procedures of measurement have been described for each case and how the authors have balanced pragmatism and accuracy in their data collection. Additionally, I consider the method of presentation (e.g. through a dashboard or through an index) as the final part of the measurement procedure, since it concerns whether and how to aggregate data.

These three steps of measurement are central to the analysis of the BWI, the SDGs and the MBW. The next section discusses multiple theoretical challenges which link to the different stages of measurement.

Challenges to the three steps: dashboards and index-making

As well-being is a concept continuously broadening in scope, multiple ways to characterize, represent and measure it have been developed. The dashboard approach and the index approach are central in this debate and both run into obstacles when it comes to the steps of measurement. For the dashboard approach, the representation of (sustainable) well-being through a variety of separately reported indicators is potentially problematic: to what extent does this set of indicators paint a picture of overall well-being? In the analysis of the SDGs and the MBW, the

extent to which these indicators represent the concept of (sustainable) well-being will be further discussed.

The index approach runs into two obstacles: 1) the axiomatic and the instrumental approach to index-making and 2) the attribution of weights to indicators. Both are problematic from the procedural perspective of the three measurement steps. The axiomatic approach to indices stipulates that the measurement procedure to derive an index number should adhere to certain predetermined rules, the axioms, in order to be valid (Reinsdorf, 2007). Importantly, these axioms cannot be inconsistent with one another, meaning that any measurement procedure not adhering to all axioms cannot be valid. The instrumental approach to index-making and measurement is more flexible: it acknowledges that certain values must be weighed and that designers of measurement procedures must make compromises (Boumans, 2001). A measurement procedure in line with this instrumental approach is likely to be imperfect but better than its alternatives.

An example of an axiom can be found in inequality measurement, namely the Pigou-Dalton principle of transfers. This principle or axiom stipulates that an inequality index must take on a lower value if a transfer of welfare from a richer to a poorer household takes place, provided that the richer household remains richer still and that total income remains the same (Hindriks and Myles, 2013). A researcher adhering to the axiomatic approach and to this specific principle cannot compromise on this when developing a measure for inequality. In the instrumental approach, practical considerations could overrule this principle when necessary, balancing theoretical requirements with an empirical reality.

As axioms for the measurement of a well-being index are at least as difficult to define as the concept itself, practice shows that researchers are forced towards the instrumental approach when developing their measurement procedure. Designers of measurement procedures must weigh a set of imperfect options and be transparent about why they make certain choices. In the end, the axiomatic and instrumental approach to index-making are exemplary of the balance between feasibility and accuracy in developing social sciences measurement procedures.

A similar obstacle to the procedure of index making is the attribution of weights to different indicators which together compose the index. This is a normative as well as a practical obstacle. In the example of measuring democracy, should a researcher combine a freedom-indicator and a fairness-indicator into one index to measure free and fair elections? Are free but unfair elections democratic? Are unfree but fair elections democratic? And which of these should be attributed more weight? In the analysis, I will return to the procedural but

fundamental obstacles to measurement and reflect on the way in which each case approaches the problem of attributing weights.

Challenges to the three steps: defining and discounting (future) well-being

The challenges of measuring concepts in social sciences are not limited to the steps of representation and procedure. For well-being and sustainability measurement, characterizing the concept comes with at least three challenges as well. Over time, different definitions of well-being and of sustainability have existed. Which definition is used determines how one can measure well-being and the extent to which one must measure and discount future well-being.

Let us first consider the definition of well-being. The utilitarian approach to well-being, developed by Bentham in 1789, maximizes the sum of everyone's existing and experienced well-being (Koopmans, 2022). Bentham's utilitarianism does not take questions of inequality into account. The Rawlsian utilitarian approach to well-being uses the 'maximin principle', stipulating that society's level of well-being is equal to the well-being of those who are worst off (Hindriks and Myles, 2013). This second approach does take inequality into account. While these two approaches have been omnipresent in economics and utility maximization, Sen (2000) proposed an alternative and influential perspective of how to define well-being. He proposed to look at development as a process of increasing human individual freedom and the capabilities of individuals. Income, health, nutrition and other forms of utility are merely means to that end (Sen, 2000).

This capabilities approach to well-being culminated in the report by Stiglitz, Sen and Fitoussi (2009) in which a person's life is seen as "a combination of various "doings and beings" (functionings) and of his or her freedom to choose among these functionings (capabilities)" (p. 42). This approach broadens the definition of well-being beyond actually experienced individual well-being. Instead, it argues that the experienced capabilities of people are a proportion of the potential and achievable capabilities which does their life justice. In this way, capabilities *not* experienced become as important as capabilities experienced. Two out of the three selected cases for this study pay tribute to Sen's approach to well-being, the BWI and the MBW. The SDGs are a successor of the HDI which also pays tribute to the capabilities approach (UNDP, 1990). Moreover, the justification of the SDGs is similar to the capabilities approach by Sen, for example when stating that the member states are determined to "ensure that all human beings can fulfil their potential in dignity and equality" (UN, 2015, p. 5). As such, Sen's capabilities approach is the philosophical foundation of these measurement methods and will be central in the analysis of how well-being is characterized for each case.

The second challenge to characterizing well-being for measurement is whether or not to incorporate sustainability and subsequently, which definition of sustainability to employ. If one decides to measure sustainable well-being, this requires a definition of sustainability. Sustainability in this debate generally concerns the capital which the current generation leaves behind for future generations, to be categorized into two schools of thought and into two respective definitions: strong and weak sustainability. There is a bundle of knowledge, goods and resources which one generation leaves behind for the next. The extent to which the current generation can substitute one type of capital in this bundle for another determines these definitions (Heath, 2016). Weak sustainability allows current generations to substitute one type of capital for another, for example leaving behind less natural resources in exchange for knowledge about how to efficiently use remaining resources. Strong sustainability does not allow such substitution. Social scientists must justify which type of sustainability they want to measure.

When one decides to value future well-being, the third challenge to well-being measurement arises: how can one predict future well-being and how does one discount future well-being to determine its current value? It goes beyond the scope of this theoretical framework to dive into the lively normative debate on discounting future well-being (see for example Drupp et al., 2018). However, a brief note is required. Economists use discount rates to compare the value of welfare now and at different points in the future by calculating the present value of future welfare. The discount factor employed can depend on expected economic growth rates, risk and a pure time preference (Heath, 2013). This pure time preference generally represents the impatience of humans in obtaining welfare now over obtaining it in the future. In case of well-being and sustainability, this time preference would represent the solidarity between generations, i.e. the extent to which current generations are willing to forego well-being for future generations.

Economists and policymakers frequently use discount rates. However, many philosophers disagree with this method, arguing that discount rates are discriminatory against future generations as their welfare is as valuable as welfare of currently living individuals (Heath, 2013). Rebuttals of this argument exist and concern the possible extinction of life on earth and the value of innovation and growth now over resources in the future (Roemer, 2011; Heath, 2013). Finally, just like these *intertemporal* distributive issues, there are *intratemporal* distributive issues concerning the division of current well-being between countries (Roemer, 2011). What type of discounting would be ethical for these issues?

The challenges above show that measurement in social sciences requires a framework. The three steps of measurement proposed by Cartwright and Runhardt (2014) are such a framework, albeit imperfect. The challenges to each of the steps discussed, show how social scientists and statisticians must make and justify choices when developing a measure. These choices are central to this research. In the next chapter, I elaborate on the methodology through which these choices are analysed for the three cases.

4. Methodology

The purpose of this study is to understand the fundamental problems of measuring (sustainable) well-being and to make recommendations towards the improvement of the *Brede Welvaartsindicator*. To gain this understanding, it is necessary to focus on a limited selection of cases. An in-depth analysis of each step of measurement as developed by Cartwright and Runhardt (2014) is required. Moreover, it is necessary to understand the deliberations made in developing these measures at each step along the way. For this study then, a multifaceted case-study analysis is the most appropriate way to answer the research questions. In this section I introduce and justify the selected cases: the main case is the *Brede Welvaartsindicator* by Utrecht University and RaboResearch and the two other cases are the Sustainable Development Goals of the UN and the *Monitor Brede Welvaart* of the Dutch statistical agency (CBS). Next, I elaborate on the ways in which case-study analysis was performed and on the methods through which data for each case was collected.

First however, why use case-study analysis in the first place? The purpose of case-study analysis fits the research questions of this study, namely to find similarities and differences between a select set of phenomena (Starr, 2014). Moreover, the exploratory nature of this research, in which data needs to be contextualized and in which each phenomenon studied contributes to theory development, makes case-based research a suitable approach (Edwards, 1998). To understand why and how the SDGs and the MBW incorporate sustainability while the BWI does not, a qualitative in-depth analysis is required. Moreover, this research is as much a study into the field of methodology as it is a study into the field of economics. To make a clear distinction: a study in methodology involves the question of *why* a certain method is used to provide information about the economy, while a study in methods involves the question of *how* to most appropriately provide information using a certain method (Boumans and Davis, 2014). Statistical analysis, field experiments and other research methods in social sciences cannot provide answers to the *why* questions of this research as well as case-study analysis.

Before turning to the multiple ways in which case-study analysis was performed, I turn to the justification of the cases.

Case study selection: the BWI, the SDGs and the MBW

A brief introduction of each case is in order. The *Brede Welvaartsindicator* was developed by Utrecht University researchers in cooperation with RaboResearch. First published in 2017, the BWI summarizes 23 indicators across 11 dimensions of well-being into one index number. Each dimension is weighted in accordance with results from subjective well-being surveys which visitors to the OECD Better Life Index fill out online (Rijpma et al., 2017).

The *Monitor Brede Welvaart* was first published in 2018 by the CBS after being commissioned to do so by the Dutch parliament. The MBW consists of the dimensions ‘here and now’, ‘later’ and ‘elsewhere’ and has contained between 53 and 73 indicators since it was published (CBS, 2018a; 2021). Multiple themes containing these indicators have been developed for each dimension, such as ‘work and schooling’ and ‘security’ for the dimension ‘here and now’ (CBS, 2018a). The CBS tracks whether there is an upward or downward trend in comparison to the values of previous years or whether an indicator is moving towards or away from a nationally or internationally determined policy target (CBS, 2019). It does not publish a weighted index.

Table 1. Description of cases

Measure	<i>Brede Welvaartsindicator</i>	<i>Sustainable Development Goals</i>	<i>Monitor Brede Welvaart</i>
Author	UU & RaboResearch	United Nations	CBS
First publication	Rijpma et al. (2017)	United Nations (2017)	CBS (2018)
Publication years	2016 to 2019	2016 to 2021	2018 to 2022 ¹
Approach	Indicator	Dashboard	Dashboard
Dimensions	11	17	3
Indicators	23	231	53

¹ For purposes of planning, the 2022 publication of the MBW has not been used in this research.

Finally, the Sustainable Development Goals were published by the United Nations in 2015 and ratified by UN member states. In total, the 17 goals are accompanied by 169 more detailed targets. In 2017, an indicator framework for the SDGs consisting of 231 indicators was published (United Nations, 2017). The UN does not publish a weighted index of the targets. Further descriptive details can be found in Table 1.

These three selected cases share similarities and have differences which are relevant to finding an answer to the research questions. All cases share a similar philosophical background, being based on or closely resembling the capabilities approach to well-being and human development which Sen (2000) proposed. Each case measures a version of well-being which arises from this background and each case uses a variety of dimensions and indicators to operationalize the concept. While these similarities make a reasonable comparison of the cases possible, the multiple differences make this comparison valuable to answer the research questions. The most notable differences are listed below:

- 1) While the *Brede Welvaartsindicator* focuses solely on well-being, the *Monitor Brede Welvaart* and Sustainable Development Goals incorporate sustainability.
- 2) All three cases have different definitions when it comes to what well-being is.
- 3) While the BWI results in a weighted index number representing well-being in line with the index approach, both the MBW and the SDGs are dashboards of indicators.
- 4) While the BWI and MBW primarily focus on Dutch well-being as object of research, the scope of the SDGs consists of all 193 UN member states which ratified the goals.
- 5) For each case, there is a different actor which initiated or commissioned the measure with different target audiences and different purposes.

The variation in measures translates to different decisions being made in each step of measurement as proposed by Cartwright and Runhardt (2014). For example, the incorporation of access to clean drinking water differs between the MBW and the SDGs: in the Netherlands, access to clean drinking water is omnipresent to such an extent that it is not worthwhile measuring. Globally however, this remains an important challenge. Dutch MBW researchers therefore chose to measure the quality of clean drinking water (de Jong, personal communication, May 31, 2022). These differences help in understanding the choices which researchers face when measuring well-being, especially when incorporating sustainability. The methods of case-study analysis are discussed next.

Case-study analysis: desk research and the typology of indicators

While case study analysis refers to an in-depth analysis of one or a couple of fitting cases, multiple qualitative and quantitative methods to perform this analysis remain. Generally, using multiple types of analysis will provide different perspectives and improve the quality of the research (Starr, 2014). For this research, I made use of desk research, I developed a typology of indicators and I conducted expert interviews. These methods complement each other. Desk research provides detailed knowledge of the methodology used for each measure of well-being. By applying a typology of indicators to the indicator set of each case, the differences between these sets becomes clear. Finally, the expert interviews provide an opportunity to ask researchers involved with the cases why certain decisions were made. A brief elaboration on each type of analysis is in order.

The set of documents subject to desktop research for each case consisted of founding documents, technical annexes to these documents, annual updates of each measure and other relevant publications. Where necessary, secondary literature was consulted to answer remaining questions and understand the place of each case within the field of well-being measurement. Through a close reading of these documents, I was able to determine which choices were made for each step of measurement by Cartwright and Runhardt (2014). For example, which definition (characterization) of well-being is used in the *Monitor Brede Welvaart*? Which indicators represent this definition and importantly, how fitting are they? Finally, is there a clear procedure of how data is collected in which way this data is reported? Analyzing these documents for each case provided valuable insights into the steps of measurement.

Aside from a general inquiry into the characterization, representation and procedures of each case, I developed a typology of the indicators which are used to measure well-being. This typology in itself is not a result of the analysis, but rather a tool through which to understand the similarities and differences between the three cases. The typology consists of three types of indicators, each being a different approach to measuring well-being. Table 2 shows this typology with the characteristics of each type and an example from the SDGs. This typology was adapted for and applied to all indicators of the BWI, the SDGs and the MBW.² As it is tailored to the well-being indicators of the selected cases, this typology is non-exhaustive. Still, it helped in finding out why certain indicators were used to measure a given dimension of well-being. Drafting the typology also helped in critically assessing why an

² Upon request, the application of this typology to each of the indicator sets is available.

indicator was chosen to represent a dimension of (sustainable) well-being. Moreover, it raised questions: Why do the authors of the MBW almost exclusively use flow indicators to represent the effects of Dutch well-being on well-being elsewhere? And why are there many policy indicators in the SDG indicator set, while there are no policy indicators in the MBW and BWI? In the analysis, I will return to these questions.

Table 2. Typology of well-being indicators

Type of indicator	Characteristics	Example from the SDGs (United Nations, 2022a)
Stock indicator	This measures the stock value of something related to a dimension of well-being, often defined as a % of total output, production or a certain population.	The proportion of agricultural area under productive and sustainable agriculture.
Flow indicator	This measures the flow of or change in a value of something related to a dimension of well-being, often defined in comparison to a predetermined standard or to previous values.	Change in water-use efficiency over time.
Policy indicator	This measures a number of geographical entities (countries, regions or municipalities) which have adopted and/or implemented a certain policy, strategy or action plan which contributes to a dimension of well-being.	Number of countries that adopt and implement local disaster risk reduction strategies in line with national disaster risk reduction strategies.

Case study analysis: expert interviews

The expert interviews supported this analysis by answering questions which were raised during the desktop research or when applying the typology. Three semi-structured expert interviews were conducted throughout the timeframe of analysis. The questions were tailored to each interviewee, but each set of questions consisted of two themes: 1) General well-being measurement and its impact on public policy and 2) incorporating sustainability into well-being: how and why? Each interview was conducted in Dutch, lasted approximately 45

minutes, and was transcribed.³ Appendix A shows an example set of questions, tailored to an interviewee closely involved with the *Monitor Brede Welvaart* through the CBS.

Together, the desktop research, the typology and the expert interviews have given a comprehensive picture of the choices made in developing the BWI, the SDGs and the MBW. These cases show the difficulty of measuring well-being, especially when incorporating sustainability. In the next section, the analysis of these choices is discussed, together with the fundamental problems of well-being measurement which have come forward in the analysis. These fundamental problems have implications for incorporating sustainability into the BWI.

5. Analysis

In this section, the results of the research are presented, analysed and translated to answers for the remaining research question. The first part of this analysis answers research questions 3, 4 and 5 (see frame below) for each of the selected cases. These answers are based on the desk research, the application of the typology of indicators and on the expert interviews where applicable. The second part of this analysis is a comparative analysis of the three selected cases. First, I discuss the societal impact of the measures, followed by the four fundamental problems of measuring well-being and sustainability. Acknowledging these problems and their imperfect solutions, I finally present recommendations to improve the *Brede Welvaartsindicator* in the third part of this analysis.

Individual cases: characterization, representation, procedure

-
3. How do the BWI, SDGs and MBW characterize (sustainable) well-being?
 4. How representative is the set of indicators chosen to measure the BWI, the SDGs and the MBW?
 5. To what extent does the procedure through which to measure the indicator fit the representation and characterization of (sustainable) well-being?
-

The publications by the author of each measure largely provide answers to how a concept was characterized, represented, and measured through appropriate procedures. For the *Brede Welvaartsindicator* and for the *Monitor Brede Welvaart*, the expert interviews and secondary literature provided answers to remaining questions. For the Sustainable Development Goals, I relied only on secondary literature. Table 3 shows the overall application of the typology of

³ Upon request, all Dutch transcripts are available.

indicators to the indicator set of each measure and the aggregate set of indicators.

Table 3. The typology of indicators applied to the cases

Measure	<i>Brede Welvaartsindicator</i>	<i>Sustainable Development Goals</i>	<i>Monitor Brede Welvaart</i>	<i>Total</i>
Indicators (N)	23	231	53	307
Stock indicators (N)	23	157	41	221
Stock indicators (%)	100	68.0	77.4	72.0
Flow indicators (N)	0	33	12	45
Flow indicators (%)	0	14.3	22.6	14.7
Policy indicators (N)	0	41	0	41
Policy indicators (%)	0	17.7	0	13.4

Brede Welvaartsindicator

The BWI is clear about which concept it measures and how it presents these results, choosing to make a practical contribution to the well-being debate. The authors choose to conceptualize and measure well-being instead of sustainable well-being and to present the results using the index-approach instead of a dashboard-approach (Rijpma et al., 2017; Badir et al., 2017). The stated purpose is to make a practical contribution to the literature on well-being. An added purpose at the time of publication was to challenge the dashboard-approach to well-being measurement: by proposing an index-approach, albeit imperfect, the authors reintroduced the approach and policymakers and other researchers were forced to relate to the BWI (van Zanden, personal communication, May 19, 2022). While the authors acknowledge the necessity of developing an indicator which incorporates both well-being and sustainability, having one indicator represent both concepts was deemed too complex in the first publication (Rijpma et al. 2017). Specifically, measuring well-being here and now is quite different from measuring the sustainability of well-being in the future. The latter requires knowledge about the future or an estimation of insecurity about the future, along with an inevitably arbitrary discount rate (van Zanden, personal communication, May 19, 2022).

Although the focus on well-being makes defining the concept to measure less complex, a clear definition is missing in the primary sources. Instead, the authors refer to Stiglitz, Sen and Fitoussi (2009), who have identified eight dimensions of well-being which should all be

considered simultaneously (p. 14-15): Material living standards, health, education, personal activities including work, political voice and governance, social connections and relationships, the environment (present and future conditions) and economic and physical insecurity. The reference to Stiglitz, Sen and Fitoussi (2009) returns throughout the publications. I consider these dimensions the characterization of well-being for this case.

In developing a set of indicators which could represent well-being as defined by Stiglitz, Sen and Fitoussi (2009), the practical purpose and pragmatism of the authors becomes noticeable. They step away from the eight dimensions of well-being. Instead, 23 indicators were chosen to represent the 11 dimensions of well-being which were developed by the OECD in 2011 (Boarini, 2011). All of the 23 indicators are stock indicators. While the OECD dimensions are similar to the eight dimensions of Stiglitz, Sen and Fitoussi (2009), they are not the same. According to Cartwright and Runhardt (2014), the indicators which represent a concept should be developed on the basis of the characterization of this concept. The discrepancies between the definitions then become problematic. The OECD stipulates that it measures both subjective and material well-being (OECD, 2022), while the BWI authors intend to measure individual capabilities and functionings (Hardeman et al., 2020).

In the BWI, there are no decision rules which justify why a certain indicator was chosen to represent well-being, aside from a general discussion on potential indicators in the initial publication (Rijpma et al. 2017). In these discussions, practical considerations are often decisive. For the education dimension for example, 'educational attainment' was preferred over 'population literacy' because the latter had become irrelevant for the developed OECD countries (Rijpma et al, 2017, p. 10). This choice was made to fit the target audience of the measure. Similarly, for the 'safety' dimension, the authors combined homicide rates with data on other violent crimes because this is the usual procedure. However, whether these indicators fit the approach by Sen, in which perceived well-being and the potential capabilities of individuals are central, is not discussed. Why not use subjective safety data to measure safety instead, for example?

While the characterization and representation of well-being are imperfectly reported in the BWI documentation, the procedure through which to make one index number out of these indicators is well-documented. Rijpma et al. (2017) describe how multiple variables within one dimension are normalized and weighed before using the subjective well-being data from the OECD Better Life Index to arrive at an index value. Although a close reading of the footnotes throughout multiple editions of the BWI is required, one can replicate the calculations given the indicator data.

This procedural rigor is what makes the BWI valuable as well as vulnerable. The added value of the BWI is that it produces one index number based on the OECD respondent's subjective valuation of the 11 well-being dimensions. The issue of attributing weights in index-making is controversial, but through this method the authors try to find an empirically supported distribution of these weights. With BWI index data from 2006 to 2020, the authors can recognize time trends and compare changes in the BWI to changes in GDP. The time trend and comparative analyses have shown that it took a decade for Dutch well-being to recover from the 2008 financial crisis (Badir et al., 2017). Similarly, the 2020 publication of BWI came with a warning on well-being trends: a crisis such as the Covid-19 pandemic could have delayed effects on Dutch well-being and take yet another decade to overcome (Aalders et al., 2020). However, for an index number to be fabricated over time, all data must be available within the requested time period. This constrained the initial authors in their data selection (Rijpma et al., 2017). Moreover, the authors had to make significant changes in the set of indicators used when developing regional versions of the BWI (Badir et al., 2017). The added value of a composite of indicators breaks down with limited data availability and in the case of the BWI, this is a vulnerability.

Sustainable Development Goals

The SDGs, adopted by the UN General Assembly in 2015, have been a landmark in climate policy history and in redefining what constitutes development. The SDGs do not necessarily define or measure well-being, but are rather an expression of what countries should strive for in developing and what constitutes a good life. Because the document is a political compromise between all 193 UN Member States, the documentation of the measure's characterization, representation and procedures are less straightforward than for the BWI or MBW. Still, between the founding document, technical annexes and other publications, a clear picture of the SDGs arose. The scope and agenda-setting power of the goals, targets and indicators have significantly impacted the debate on well-being measurement. For example, recent MBW publications report on the Dutch progress for all 17 goals and the SDGs even give direction to the interpretations by the MBW authors, the CBS (de Jong, personal communication, May 31, 2022).

Sustainable development is characterized as development that fits the Brundtland definition of sustainability (World Commission on Environment and Development, 1987, p. 41): "sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs", with development

being "a progressive transformation of economy and society". For the United Nations (2022c, June 12) and for the Sustainable Development Agenda, this is still the guiding definition.

While on the one hand this is a clear definition of what sustainable development is, this characterization allows for subjective interpretations and raises questions about the UN priorities. What constitutes a progressive transformation of society? Which future generations must we consider? Mair et al. (2018) argue that this definition and the 17 goals are highly contested and politically sensitive. The political sensitivity of this definition arguably is a result of the political negotiations which preceded the SDG publication. Dang and Serajuddin (2020) point out that there is no clarity about the priorities between achieving national and global goals and between goals which conflict with one another. Consider a country which simultaneously shows poverty reduction and an increase in inequality, respectively goal 1 and the opposite of goal 10. Is it developing sustainably? The founding document of the SDGs merely states that the goals are "integrated and indivisible" (United Nations, 2015, p. 3).

The indicators which represent 'sustainable development' are indicative of the choices *not* made at the SDG establishment. They showcase the agenda-setting nature of the project. The global indicator framework was only agreed upon in 2017, two years after the establishment of the SDGs. An Inter-Agency and Expert Group on SDG indicators (hereafter: IAEG-SDGs) was assigned to develop the indicator framework which was adopted by the General Assembly of the UN in June of 2017 (United Nations, 2022b, June 12). The responsible statisticians faced multiple obstacles in determining the framework. Many politically sensitive concepts for these goals were not defined clearly and statisticians were asked to find just one representative indicator for each of the 169 targets accompanying the 17 goals (Macfeely, 2020). Essentially, while the characterization of sustainable development should help to deal with the 'fuzziness' of such a contested concept (Cartwright and Runhardt, 2014), many questions remained for the statisticians.

There are both upsides and downsides to the remaining 'fuzziness' of the concept 'sustainable development'. On the one hand, there are indicators which do not accurately measure the target which they should measure, which is problematic because these indicators are shaping policy (Mair et al, 2018). On the other hand, while indicators are limited in the extent to which they can represent a certain concept, they can become an anchor for debate and critical reflection on these concepts (Mair et al, 2018). For example, target 16.6 is to develop "effective, accountable and transparent institutions at all levels", which is measured by two indicators: "government expenditures as a proportion of original approved budget" and the "proportion of the population satisfied with their last experience of public services" (United

Nations 2022a, p. 20). While these indicators are imperfect in measuring how effective or transparent an institution is, they do start a debate on how to measure the quality of institutions in the first place. This gives the SDGs and specifically the Global Indicator Framework agenda-setting powers in the field of statistics.

Applying the typology of indicators to the SDG indicator set showcases the agenda-setting nature of the indicator framework for national and local governments. Policy indicators are extensively used. The SDG indicator set consists of 157 stock indicators, 33 flow indicators, and 41 policy indicators. The policy indicators keep track of whether or not policies are developed or implemented at a certain level of government. While a policy indicator does not measure the actual progress made towards a certain target, it does force (local) governments to make or implement policy striving towards this goal. An indicator exemplifying this is indicator 14.c.1, which is coupled to the target of enhancing and conserving the sustainable use of oceans (United Nations, 2022a, p. 17):

[The] number of countries making progress in ratifying, accepting and implementing through legal, policy and institutional frameworks, ocean-related instruments that implement international law, as reflected in the United Nations Convention on the Law of the Sea, for the conservation and sustainable use of the oceans and their resources.

While such an operationalization risks governments drawing up a plan without implementing it, this does put conservation of oceans on the agenda. Besides, it provides national politicians who value an SDG target with a new tool to control their government's actions towards meeting its international obligations. The way in which sustainable development is represented in the SDGs, especially through policy indicators, is therefore setting local and national policy agendas around the world.

The third step of measuring the SDGs, the procedure, shows the challenges which current statisticians face around the world. Many indicators, even after being defined, required and still require the development of measurement procedures which can be adopted by statistical offices around the world. Dang and Serajuddin (2020) show that most data are still missing, with only 19% of all required data to track the goals available. The IAEG-SDGs has ranked indicators into three tiers based on the methodological development and data availability (UNSD, 2022, June 13): tier 1 indicators are conceptually clear, operationalized and have data reported for at least 50% of countries and of the population for which this indicator is relevant; tier 2 indicators are conceptually clear but have limited data availability;

tier 3 has no established methodology available yet. SDG target 11.C, regarding support for developed countries in constructing sustainable and resilient buildings, does not even have an agreed upon indicator yet (United Nations, 2022a). Essentially, the complexity of measuring the SDGs is a result of the struggles and the ambitions of the politicians who proposed the goals: Fuzzy targets have been difficult to accurately measure, while ambitious targets have required the development of new types of measurement on the other hand. In light of this, it is not surprising that the final two targets of the 17th goal are concerned with statistical capacity building, progress in SDG measurement and the aggregation of timely, high-quality data (United Nations, 2022a).

Monitor Brede Welvaart

Like the BWI, the MBW focuses on Dutch well-being. The MBW was developed by the Dutch central statistical agency (CBS) and commissioned by the Dutch government. The initial proposal however came from a temporary committee of the Dutch parliament (*Tweede Kamer*). The members of parliament advised the government to further develop an existing program into the *Monitor Brede Welvaart*, noting that it should be an internationally comparable and visually attractive dashboard of information (Grashoff et al., 2016). The government has adopted this advice and the MBW is discussed annually in May, when the Dutch finance minister must account for the previous year's budget. The purpose of the MBW is to provide a factual basis for the political and societal debate about well-being and the MBW has been developed with this neutrality in mind (CBS, 2018; de Jong personal communication, May 31, 2022). The authors of the MBW have adopted the dashboard approach because they do not deem it possible to combine a diverse set of well-being aspects into one indicator (CBS, 2018). In all three steps of measurement, the CBS is transparent about its choices and how these choices were made, balancing the quality of indicators and the required data availability.

The characterization of *brede welvaart* (comprehensive well-being) is made explicit in the first CBS publication and is linked to Sen's capabilities approach and the Brundtland definition of sustainable development. Well-being for the MBW is "the quality of life here and now and the extent to which this is at the expense of welfare of later generations or current generations elsewhere in the world" (CBS, 2018, p. 18, author's translation). This definition of well-being is divided into three dimensions: 'here and now', 'later' and 'elsewhere' (UNECE, 2013). These three dimensions are in line with international recommendations by the Conference of European Statisticians (CES), which in turn follows the Brundtland definition of sustainable well-being and the recommendations by Stiglitz et al. (UNECE, 2013). These

three dimensions are all a part of well-being when viewed comprehensively and are conceptually different to such an extent that they must be separated in measurement (de Jong, personal communication, May 31, 2022). Sen's capabilities approach is a pillar of the CES recommendations and the concepts 'capabilities', 'functionings' and 'values' are linked to the types of indicators selected for the MBW: input, output and outcome (CBS, 2018b). The CBS is transparent about its definition of well-being, its client, its purpose and its philosophical background.

In the second step, the representation of well-being through indicators, the authors are again transparent. To minimize the influence of the authors' personal preferences and intuition, the CBS has developed a strict system with decision rules to select indicators and eventually present the results (CBS, 2018b). Every indicator of the MBW measures a theme from the CES-recommendations and is selected based on six factors of data quality, such as validity, the reliability of the source, the internal consistency and the extent to which the indicator can be compared to other countries (CBS, 2018b). This has resulted in a set of 53 indicators in the first edition, with 41 stock indicators, 12 flow indicators and 0 policy indicators.

What stands out is that 11 out of 12 indicators which measure well-being 'elsewhere' are flow indicators. These indicators *inter alia* measure the depletion of natural capital such as metals elsewhere due to imports to the Netherlands. Official development assistance and other transfers are measured as a benefit to people elsewhere. Through these flow indicators, the CBS essentially measures the well-being transfers of trade with other countries, making a distinction between least developed countries and other countries (CBS, 2018). The flow indicators are useful in representing both the benefits and drawbacks of international trade (de Jong, personal communication, May 31, 2022). However, they miss other ways in which Dutch well-being affects well-being abroad (Stoel, personal communication, May 20, 2022). The Dutch tax system could for example affect foreign tax income through its corporate tax rate. Similarly, Stoel proposed that migration of highly educated people to the Netherlands could affect human capital abroad (personal communication, May 20, 2022). The authors of the CBS acknowledge this limit of the MBW. Nevertheless, they argue that the focus on well-being 'elsewhere' is quite a novel approach and that there are no suitable CES-recommended indicators yet for many capital flows between countries (CBS, 2018).

The MBW shows that finding the right indicators to represent a dimension of well-being is difficult. The CBS is not deaf to critique of its method from politicians and other external commentators, but not every dimension of well-being is easily captured in an indicator. De Jong, a CBS developer of the MBW, argued that both migration and culture were

underrepresented in the measure, but both require good indicators which are still to be found (personal communications, May 31, 2022). De Jong used the examples of barn fires to show how difficult this can be. When researching how to include barn fire data to represent animal suffering, the CBS found that both the annual number of barn fires and the annual number of animals killed in a barn fire could not represent this concept (de Jong, personal communication, May 31, 2022). A chicken barn fire causes many more animal casualties than a kettle barn fire, making aggregate data very volatile to what type of barns burned down in a given year: A year with many barn fires might see relatively few animal casualties, while a year with few barn fires could have many. This made it impossible for the CBS to distinguish a trend over time regarding animal safety. Data is only useful to the MBW if the CBS can distinguish trends and this example shows that not every indicator fits the purpose. Still, the CBS continues to develop and improve the indicators matched to each dimension of well-being. The number of core indicators has increased from 53 to 73 between 2018 and 2021 (CBS, 2018; 2021).

The CBS elaborates on the final step of measurement, the procedure, in an additionally published explanation to the MBW. This document contains the decision rules used to select indicators, the statistical methods used to determine trends and the decision rules used to determine how trends are labelled in the dashboard (CBS, 2018b). Statistically significant increasing or decreasing trends are labelled green or red, insignificant trends are labelled grey. The CBS (2018b) claims that the procedure explained in the additional publication theoretically allows others to replicate the results and even the publication of the MBW itself. To be as objective as possible, the CBS report on well-being is almost "boring" and only points out the most obvious relations: The correlation between the decrease in people's satisfaction of their free time and the increase in working hours per capita may be reported, but there is no mention of causal relations (de Jong, personal communication, May 31, 2022). While the CBS explicitly does not pass judgment on whether a trend is good or bad, it does link the MBW trends to the Sustainable Development Goals. Following up on a request by the government, the CBS publishes how trends are linked positively or negatively to achieving certain SDGs (CBS, 2019). Through the SDG indicators, the CBS passes judgment on whether or not trends are in line with externally determined values, while still protecting its own objectivity as much as possible (de Jong, personal communication, May 31, 2022).

Comparative analysis: Real-world effects and fundamental problems

Having discussed the three steps of measurement for the BWI, the SDGs and the MBW, I now turn to a comparative analysis of these three measures. This analysis consists of two parts: 1) a

brief comparative discussion of the real-world impact which these measures have had and 2) a discussion of the fundamental problems of well-being measurement which arise from the previous section. Researchers developing a measure of (sustainable) well-being must make choices on the basis of their aspired impact and the fundamental problems which they face. This section and the final section containing recommendations for the BWI together answer the final research question:

6. How do the lessons learned from the SDGs and MBW apply to the incorporation of sustainability into the Brede Welvaart Indicator?

Making an impact

Which type of impact a measure makes depends on what it measures, on how this is measured and on how the results are reported. The BWI, the SDGs and the MBW have all made an impact on the world, either by starting a debate, influencing policy or setting the agenda.

The purpose of the BWI authors was to publish an index measure of comprehensive well-being and draw attention to the debate on what well-being constitutes and how it should be measured (van Zanden, personal communication, May 19, 2022). The first publication succeeded in drawing popular attention (RTL, 2017; NOS, 2017) and political attention (Provinciale Staten van Noord-Brabant, 2017). The authors also had the ambition to challenge other economists and statisticians to do better. Recent debates on how to measure well-being specifically mention the BWI and the MBW before proposing or disapproving of new initiatives, showing that the authors achieved this goal as well (Jacobs, 2021; Vollebergh, 2022). Moreover, with regional analyses and ongoing projects, the BWI remains relevant (van Zanden, personal communication, May 19, 2022).

While the MBW has also influenced the popular and academic debate in the Netherlands (e.g. Dorlo, 2018; Hoekstra, 2021), the main channel of impact has been political. The CBS enters into a conversation with Dutch politicians at least yearly when the government accounts for its budget, and the tone in those conversations has shifted: where the first conversations in 2018 revolved around explaining what the MBW is and how it works, more recent conversations with the legislative branch were about the role with this measure could play in all phases of the Dutch policy cycle (de Jong, personal communication, May 31, 2022). Meanwhile, multiple ministries in the executive branch of government are using the CBS data to monitor relevant developments in their field (Stoel, personal communication, May 20, 2022). Simultaneously, policy makers such as Stoel are trying to further improve the use of the MBW

in policymaking and looking at the role of well-being and economic growth in strategy-making (personal communication, May 20, 2022). The desktop research and expert interviews jointly paint a picture of cooperation between the CBS and the Dutch government to make the best use of the MBW in policymaking.

Although the SDGs play a role similar to that of the MBW and the BWI at the global level, the set-up of the indicators and the scale of the operation cause its impact to be different. The policy indicators which are included in the SDGs cause statistical agencies and governments to pay attention to sectors to which no attention has been paid before. This agenda-setting nature can have a positive, but also a negative impact. Focusing on one dimension can damage efforts in another. Mair et al. (2017) point out that while the right SDG indicator may for example prioritize policies on aids prevention, the lack of indicators which involve cultural well-being may cause a reduction in efforts which promote this dimension. Ittis and Matthews (2017) argue that even acknowledged challenges like the eradication of neglected tropical diseases are benefited by the agenda-setting power of the SDGs, as this challenge competes with other challenges for limited resources and attention. Experts recognize that the SDGs have more agenda-setting power than the BWI or MBW does (Stoel, personal communication, May 20, 2022; de Jong, personal communication, May 31, 2022). One expert added that the SDGs have impact outside governmental policy as well, with businesses making commitments in support of the goals (de Jong, personal communication, May 31, 2022). In comparison, for both the BWI and the MBW, no evidence of businesses committing was found.

What type of impact researchers want to make in the societal, academic or political debate should partially guide their decisions when developing a measure for well-being. The three cases in this study are examples of this dynamic. More important still is how researchers deal with the fundamental problems that come with (sustainable) well-being measurement, to which I turn now.

Four fundamental problems of well-being and sustainability measurement

Measures of (sustainable) well-being are imperfect: indicator sets are limited in the scope of what they can cover and in the extent to which they accurately represent the concept. From the desk research, the application of the typology and the expert interviews, I have derived four fundamental problems which researchers face when measuring well-being. Each problem is linked to at least one of the three steps of measurement which Cartwright and Runhardt (2014) developed.

Problem 1. The problem of defining a living concept

The 'fuzziness' of concepts such as sustainability, well-being and development make it difficult to characterize them for measurement. As people change and society changes, these concepts are likely to be understood differently. In measurement, there is a balance to be held between remaining timely and remaining stable. Remaining timely requires regular updates of the characterization of a concept and its subsequent representation and procedure. The downside of this is that data are not entirely comparable over time. Remaining rigorous requires researchers not to make changes to the characterization of a concept and subsequent measurement procedures. The downside of this decision is that it can reduce the relevance of a measure as people develop different understandings of the concept.

For the MBW, the authors chose to keep the three dimensions constant and only add or remove indicators if they passed the decision rules. Moreover, the MBW is accompanied by additional publications about timely challenges to (sustainable) well-being, such as resilience to external shocks and natural capital stocks (CBS, 2022; 2022b). As the SDGs were part of a political process, the 17 goals which express sustainable development are unlikely to change, just like the MBW dimensions.

Still, while the SDGs and the MBW both pay tribute to the Brundtland definition of sustainable development and to Sen's capabilities approach, their characterizations of what constitutes well-being differ: the MBW takes into account well-being 'elsewhere', whereas the SDGs do not. This dimension is not applicable to the SDGs as the UN aspires to a global approach. It shows that the target audience of a measure plays a role in determining its characterization as well. Each case at least balances between relevance and stability when characterizing a concept, making the liveliness of the concepts well-being and sustainability a fundamental problem. The target audience of a measure adds to this liveliness.

Problem 2. The problem of accuracy

Like well-being measurement in general, finding an accurate indicator to represent an aspect or dimension of well-being is necessarily imperfect. Indicators boil a dimension down to its core and in doing so, they "rely on the magic of numbers and the appearance of certainty and objectivity that they convey" (Merry, 2011, p. S84). The BWI and the MBW represent one approach to finding the right indicator. These cases use existing indicators and test whether they adhere to certain decision rules (CBS, 2018b), or whether a better indicator is available (Rijpma et al., 2017). This requires compromises, with data selected more for practical reasons such as availability and reliability than for methodological reasons.

The cases show how researchers have different approaches to this problem. The indicator sets of the BWI and the MBW use previously existing indices which measure the stock levels of a type of well-being. For example, both cases use the Living Planet Index to measure the stock of biodiversity based on the population trends of 4911 species (Loh et al., 2005). This index has been in existence since 2005 and is used by the BWI authors because it is consistent, well-known and because there is a Dutch dataset (Rijpma et al., 2017).

With the SDGs, the UN has largely taken a different approach, developing many new statistical indicators for which data and data collection methods are not available yet (Macfeely, 2020). As a result though, Macfeely (2020) shows that progress on many of the targets set in 2015 is still underreported. These indicators can be more accurate than existing indicators or the indices used by the MBW and BWI, but limited data availability constrains their usefulness.

This problem is fundamental because economic methodology and statistics will continue to innovate its methods of data collection and will continue to develop new and improved indicators. Researchers will remain caught between the accuracy of their indicator set on the one hand and the feasibility and affordability of collecting the required data on the other hand.

Problem 3. The problem of future preferences and capabilities

The MBW and the SDGs, which incorporate sustainability into the measurement approaches for well-being, run into the problem that nobody can predict the future. Whether and how to incorporate future well-being into measurement is a problem of characterization as well as representation. There is a fundamental uncertainty to what the future holds, both in what constitutes well-being for future generations and in what these generations can achieve when given a certain set of capital. Multiple current debates revolve around which behaviour in the present is ethical given the uncertainty about the future. The debate surrounding narrow or broad definitions of sustainability boils down to what type of capital we can ethically leave behind. Similarly, the debate about how to discount future well-being to its present value boils down to our solidarity with future generations and how much we value their well-being. Recent discussions of these topics show that these debates remain unresolved (e.g. Heath 2016; Braithwaite and Roberts, 2021; Wagner et al., 2021).

While the debate is unresolved, each measure of well-being must make choices about the valuation of future well-being. The authors of the BWI focus on well-being of current generations, arguing that it is currently too complex to incorporate well-being of current and future generations into one index (Rijpma et al., 2017). The CBS with its MBW does take

future well-being into account but separates it from the measurement of current well-being. Even then, the current MBW approach to measuring future well-being is to determine the future value of the current capital stock with the current usefulness of this capital - the CBS explicitly does not make scenarios incorporating innovation or new types of scarcity in the future (de Jong, personal communication, May 31, 2022). With this method, the agency circumvents the use of discount rates. The UN also uses stock indicators to make an inventory of current capital, such as the proportion of marine and coastal areas which are labelled as protected areas (UN, 2017). It adds to this inventory of current capital through policy indicators which must measure progress towards sustainable use of this capital. In the end, none of the three cases uses discount rates to value the well-being of future generations, each arguably circumventing this fundamental problem.

Problem 4. The problem of presentation

The final problem of well-being measurement is in the procedure of measurement and specifically in how to present the results. The dominant debate here is between dashboards and indices and it is fuelled by the notion that there is no objective attribution of weights to different aspects of well-being. Either a researcher makes one index out of all indicators and attributes an imperfect weight to each indicator, or a researcher presents all indicators in a dashboard of information. In essence, all indicators are actually weighted equally in case of the latter. While the weights argument is central to the argument against indices, the way in which indicators are scaled is of significant influence to the final index as well (van Zanden, personal communication, May 19, 2022). The authors of the BWI show that the scale which they use to normalize indicators before aggregating them has more impact on the index than the attribution of different weights (Rijpma et al., 2017).

There is value to both the dashboard and the index approach. The index approach can be easily understood by people and allows researchers to compare well-being trends with other trends, such as in GDP growth. However, positive developments in one indicator can cancel out negative developments in another, potentially keeping an index stable while the nature of people's well-being changes. Dashboards prevent such trade-offs from taking place. By presenting all data, dashboards present a more comprehensive, albeit more complex, picture of well-being. Moreover, dashboards can point to an actor responsible for certain well-being trends: if the value of human capital is decreasing in a country, its ministry of education can be held accountable. For a well-being index, this is not the case. Or, in the words of a policy officer for the Dutch ministry of economic affairs, who commented on the publication of one number

for well-being: "Who would be responsible for the 8 or the 7?" (Stoel, personal communication, May 20, 2022, author's translation). Dashboards do however permit people, especially politicians, to cherry pick the trends which they find important, which can obfuscate the societal debate.

Researchers deal with this problem as best they can. The BWI makes use of the weights attributed to dimensions of well-being by visitors of the OECD Better Life Index website. This prevents researchers from attributing their own subjective weights. The MBW publishes interpretations of the dashboard data based on a set of decision rules and in relation to the SDGs. In both cases the authors try to find the middle ground and circumvent using their own subjective experiences. The problem however remains. An index presents a clear common trend but obscures the complexity of what constitutes well-being. A dashboard presents data on all dimensions of well-being simultaneously but blurs the overall picture. In the next section, I argue that the best approach for the BWI is somewhere in the middle.

Improvements to the BWI: increasing transparency and incorporating sustainability

In this final section, I recommend potential improvements to the *Brede Welvaartsindicator*, based on the SDG and MBW analyses and the identified fundamental problems of well-being. The goal of the BWI was to provide a practical addition to the well-being debate using an index-approach. The four recommendations have been developed with these goals in mind. Advising the authors to turn completely to a dashboard-approach would be fruitless in light of these goals. The recommendations are given below:

-
1. Shift away from setting well-being on the political and academic agenda towards improving the measure itself.
 2. Be more transparent about the definition of *brede welvaart* and the indicator selection process.
 3. increase the set of indicators which represents well-being.
 4. Report capital stock trends alongside the BWI as a representative of sustainability.
-

Recommendation 1

An important reason to publish the BWI was that at the time of publication, no index-measure of well-being existed yet (van Zanden, personal communication, May 19, 2022). The BWI publication and the MBW of the CBS have helped set well-being on the political and societal

agenda and sparked an academic debate around how to measure well-being. Nowadays, the concept is part and parcel of many political debates and initiatives, not only at the national level but also in municipalities (Smit, 2022). Of course, consulting politicians and publishing the BWI remains important to keep society engaged. However, the omnipresence of well-being in Dutch political discourse allows the authors to focus further on improving the measure itself. Van Bavel, Hardeman and Rijpma (2019) already propose five steps to improve the measure. While each proposed improvement is of value, I recommend focusing on two of these improvements. Moreover, I explicitly recommend an improvement which the authors choose not to incorporate.

Recommendation 2

Van Bavel, Hardeman and Rijpma (2019) propose finding a better connection between the measurement procedure and the capability approach. The analysis of the BWI through the lens of the three steps of measurement by Cartwright and Runhardt (2014) shows that this is indeed an underdeveloped part of the BWI. One way to improve on this is to use surveys to improve knowledge of peoples experienced capabilities (van Bavel, Hardeman and Rijpma, 2019). While such data is valuable, I propose focusing on the characterization and representation of *brede welvaart* first. A clear definition of what constitutes *brede welvaart* and how this relates to the capabilities approach is missing from the BWI publications. As *brede welvaart* is a living concept subject to the first of the four identified fundamental problems, having a clear in-text definition is important.

Similarly, Rijpma et al. (2017) explain why an indicator fits a dimension of well-being better than other indicators, but an explanation of *how* this indicator fits the capabilities approach is missing. It would be valuable for readers to understand the link between an indicator and the dimension it represents better. The decision rules used and published for the MBW could serve as an example of how to transparently select and justify the indicator set. By assessing which indicators are poorly suited to represent a dimension through the capabilities approach, the authors know better which indicators to change and which data to add, possibly through the use of surveys. For example, the dimension 'health' is represented by the indicator 'life expectancy' (Rijpma et al., 2017; Aalders et al., 2019). In comparison, the MBW makes use of 'healthy years of living at birth' and the SDGs make use of a set of indicators relating to physical health, mental health, health insurance and eradication of diseases (CBS, 2018; UN, 2022a). The capabilities approach is about giving individuals the ability to live as they would like (Sen, 2000). In that sense, quality of living (i.e. being of sound

mind and health) is more important than the length of a life. Surveys could address the subjective and experienced health of Dutch people and replace life expectancy as an indicator. It would be valuable to make such an evaluation for each indicator, thereby making an inventory of indicators which are still suitable and which need improving or replacing. Through this, the BWI is made more resilient in light of the second identified fundamental problem, the problem of accuracy.

Recommendation 3

Van Bavel, Hardeman and Rijpma (2019) propose to strengthen and extend the set of indicators which measures *brede welvaart*. The BWI started with 23 indicators (Rijpma et al., 2017) and for reasons of regional data availability decreased this to 21 indicators (Aalders et al., 2019).⁴ A comparison with the MBW and SDGs, which started out with 53 and 231 indicators respectively, shows that this is indeed an important step to take. Because the BWI is more in line with the MBW approach to indicators than with the SDG approach to indicators, I recommend assessing which dimensions require further elaboration on the basis of the MBW indicator set.

One example of an indicator which fits the capabilities approach and would be a valuable addition to the BWI indicator set is the '% of people satisfied with free time' for the BWI dimension 'jobs' (CBS, 2018a, p. 36). Another valuable theme would be to include 'contact with friends, family and colleagues', represented in the MBW as '% of people over 15 with multiple monthly social interactions' (CBS, 2018a, p. 37). Of course, the MBW indicators are incomplete just like the BWI indicators are. Moreover, changing the set of indicators makes comparing *brede welvaart* over time and recognizing trends more difficult. Still, a larger set of indicators could increase the accuracy of the BWI, and with it, its relevance.

Recommendation 4

The authors of the BWI acknowledge the necessity of incorporating sustainability into the index (Rijpma et al. 2017). Van Bavel, Hardeman and Rijpma (2019) however reaffirm the focus on current well-being, stipulating that it is too complex to discount future well-being into an index of current well-being. I argue that the latter does not exclude the former and that incorporating sustainability into the BWI publications is an important step in increasing the

⁴ I consider the six *World Bank Governance* Indicators as separate indicators instead of as one. If I were to consider these indicators as an aggregate index, the decrease in variables would be from 18 to 16.

relevance of the measure. The MBW and the SDGs both exemplify, albeit imperfectly, how to incorporate not only current, but also future well-being.

The authors of the BWI are right when it comes to the complexity of aggregating current and future well-being into one index. Discounting future well-being to its current value is a contested practice, laden with ethical considerations (Heath, 2013; 2016). Moreover, what counts as an increase of current well-being can have detrimental effects on future well-being. The identified fundamental problem of future preferences and capabilities remains as current statisticians cannot predict or estimate what constitutes well-being for an undetermined number of future generations.

Instead, I recommend the authors to take an approach similar to that of the SDGs and MBW, using stock indicators to track changes in the natural, social and human capital. Both the UN and the SDGs do not discount future well-being but rather make an inventory of the current capital stock and of whether this generation leaves behind more or less of it. Whether or not to attach a value judgment to these trends is up to the authors: the UN has stated some specific goals for the development of capital stock, while the MBW just reports whether the development is up or down. Examples of such capital stock indicators are total household debt (economic capital), fossil fuel reserves per capita (natural capital) and the labour force as a proportion of the total population (CBS, 2018a, pp. 42-43).

An important caveat is that these capital stocks must be reported alongside the index of current well-being, as aggregation of both with current methods is conceptually difficult. This recommendation seemingly contradicts the index-approach to well-being measurement. However, the capital stocks make it possible to compare current well-being to what this means for the resources of next generations, increasing the value of the current indicator. The index value of the BWI has been used extensively to be compared to GDP per capita trends (Rijpma et al., 2017; Badir et al. 2017; Aalders et al., 2019). A comparison between BWI trends and capital stock trends could be equally valuable at least. Optionally, the capital stock trends could be disaggregated into natural, social and human capital before comparing it to BWI values over time. Reporting on other trends than the main index is no novelty for BWI publications. The BWI authors understandably make use of dimension-level trends and data to explain larger trends in the index.

An index consists of multiple weighted indicators and while the index itself might be a useful summary of the general trend, there is no reason *not* to present the underlying data (van Zanden, personal communication, May 19, 2022). The same is the case for the capital stocks: the index itself can remain the most important feature of the BWI publications, but there is no

reason *not* to include additional information which to some extent represents the sustainability of well-being.

6. Concluding remarks

The main research question of this study was: *To what extent can measurement of well-being incorporate sustainability?* By answering this question on a case-study basis, this research has been an attempt to contribute to both the societal and academic debate on (sustainable) well-being. Through desktop research, a typology of indicators and expert interviews, I answered the six subquestions on three different measures of (sustainable) well-being: the *Brede Welvaartsindicator*, the *Monitor Brede Welvaart* and the Sustainable Development Goals. The three steps of measurement in social sciences provided a useful framework to determine 1) which concepts the authors of the cases wanted to measure; 2) which indicators represented these concepts and 3) which procedures were used to measure and present these indicators (Cartwright and Runhardt, 2014). Through a comparative analysis, I identified four fundamental problems in measuring (sustainable) well-being. Finally, I translated these four fundamental problems to practical and feasible recommendations for the authors of the *Brede Welvaartsindicator*.

The research made clear that there is no perfect measure of well-being. Academic and societal debate remains on what constitutes well-being, which role sustainability could and should play in measuring well-being and about how to present the results. All analysed cases are flawed in their own way. The index-approach of the *Brede Welvaartsindicator* allows the authors to determine trends over time, but they are constrained by data availability. Moreover, one index number cannot do justice to the underlying developments in each dimension of well-being. The Sustainable Development Goals are ambitious, are recognized globally and are setting the agenda for statisticians and policymakers around the world. However, many indicators are insufficiently developed and data is widely unavailable. Finally, the *Monitor Brede Welvaart* is a transparent measure which performs well on the three steps of measurement and is most replicable of the three cases. However, the neutral position of the CBS in the political debate requires it to be very specific and careful in determining trends. The authors must be even more careful in determining causal relationships, limiting the explanatory power of the measure. Moreover, the MBW publications are vulnerable to cherry picking by politicians.

The analysis points towards four fundamental problems in measuring (sustainable) well-being: the problem of defining a living concept, the problem of accuracy, the problem of

future preferences and capabilities and the problem of presentation. Designers of new measures for well-being must make choices to address these problems. Rather than wanting to make a 'right' choice, designers should focus on justifying why a choice fits the purpose of the (sustainable) well-being measure best. These fundamental problems and the weaknesses in the *Brede Welvaartsindicator* have led me to make four recommendations to improve this measure: 1) move away from setting well-being on the agenda and focus on improving the measure itself instead; 2) increase transparency about which concept is measured and why the indicators fit this concept; 3) increase the set of indicators and 4) report capital stock indicators as a representative for sustainability next to the BWI itself.

A reflection on this research

This research is just as imperfect as the three cases analysed in it. A few notes on the scope, the limits and the implications of this study are therefore in order.

First, while the answers to the research questions have been carefully formulated and justified, it is unlikely that they are exhaustive. The case-study based analysis was required to dive into specific procedures of measurement, but made a comprehensive study of more well-being measures impossible. For future research, it would be valuable to extend this analysis to other types of well-being measurement. This research could test the currently identified fundamental problems and find new problems. Similarly, it could test and extend the typology of indicators. Potential objects of research could be the Human Development Index, the UN System of Environmental-Economic Accounting and the OECD How's Life initiative.

Second, sustainability is not the only concept which is difficult to measure in relation to well-being. The well-being effects of migration have been underrepresented in the MBW for example. Cultural aspects of well-being are underdeveloped as well. These cultural aspects can encompass the value people attribute to their religious life or to their ideological freedom, but could also incorporate the value attributed to visiting a theatre or taking part in other cultural events. It would be valuable to develop indicators measuring this part of well-being, in recognition of the value which it represents to people. The multidimensional poverty index, which was developed for Nepal and which includes indicators for cultural inclusion, could be a good point of departure (Wagle, 2005).

Third, the *Brede Welvaartsindicator* is the main case of this study, but the conclusions and implications extend beyond this measure. As this research departed from the *Brede Welvaartsindicator*, returning to it in the final recommendations was a logical step. However, just like the identified fundamental problems, the four recommendations are limited to what I

could conclude from the selected cases. Analysing other measures of well-being, especially those with an index-approach, could yield different or even opposing recommendations for the BWI. Similarly, the fundamental problems can be translated to improvements on the MBW and the SDGs. The first BWI recommendation on improving the overall quality of the measure, is valid for the SDGs as well. For the MBW, the recommendation to increase the indicator set could apply.

Finally, this research would be incomplete without reflecting on the subjectivity of the researcher. Like the authors of the MBW, I developed a framework within which to study these concepts. The goal of this framework was to minimize the extent to which my subjective experiences and opinions become part of the study. Of course, I am not infallible. The selected topic for example, being the relationship between measuring well-being and measuring sustainability, is of personal interest. While these concepts have become more dominant in public discourse over the past decades, there are many people for whom sustainability and well-being remain inferior to happiness or economic growth.

Despite these limits, the conclusions of this study can be of importance to academia and society. When academics develop new measures for (sustainable) well-being, the four identified fundamental problems can guide their decision-making. When the authors of the BWI recalibrate their methods, the four recommendations could provide inspiration. And finally, when society calls for an evidence-based approach to improving well-being or speeding up the transition to sustainability, this study shows that finding the evidence is harder than it looks.

Appendix A. Semi-structured interview questions (example)

These questions were sent to the interviewee beforehand and were the foundation of the interview of May 31st, 2022.

General Questions

- What type of job do you have and how are you involved with the *Monitor Brede Welvaart*?
- Within academic literature, a debate exists between the ‘dashboard’-approach and the ‘index’-approach to measuring well-being. What have been the main motivations to choose a ‘dashboard’-approach to well-being measurement for the *Monitor Brede Welvaart*?
- By not attributing any weights to indicators of the *Monitor Brede Welvaart*, a central problem in the debate on dashboards and indices is circumvented. Meanwhile however, every indicator and dimensions are respectively weighted equally in absence of other weights. How do you and your colleagues reflect on this?
- Over the last couple of years, the *Monitor Brede Welvaart* has slightly been adapted with each consecutive edition. Which considerations are at the foundation of these adaptations (e.g. adding new indicators to one of the well-being dimensions).
- Which effects do you see in the behaviour of politicians and policymakers as a consequence of the annual publication of the *Monitor Brede Welvaart*?
- Do you think it is possible to objectively measure a concept such as well-being?

Well-being and sustainability

- To what extent do current procedures to measure sustainability in the *Monitor Brede Welvaart* suffice?
- Which obstacles did you and your colleagues encounter in measuring well-being for the dimension ‘later’? For example, how do you and your colleagues account for insecurity about the future?
- Our understanding of well-being changes over time. How can instruments to measure well-being, such as the *Monitor Brede Welvaart* and the *Brede Welvaartsindicator* remain meaningful for policymakers under these changing circumstances?
- The *Monitor Brede Welvaart* is currently linked to the Sustainable Development Goals. What is the added value of the *Monitor Brede Welvaart* in relation to the Sustainable Development Goals? In other words: how do these measures differ in your expert opinion?

Bibliography

- Aalders, R., Bavel, B. van, Hardeman, S., Lippe, T. van der, Philips, R., Raspe, O., Rijpma, A. and Stam, E. (2020). Laat herstel brede welvaart niet weer tien jaar duren. Retrieved from: https://www.uu.nl/sites/default/files/190620_bwi_def.pdf
- Aanhangsel Handelingen II, nr. 81, item 2.* (May 18, 2022). Tweedekamer.nl. Retrieved from: https://www.tweedekamer.nl/kamerstukken/plenaire_verslagen/detail/2021-2022/81#id3953152a
- Badir, M., Bavel, B. van, Hardeman, S. and Rijpma, A. (2017). Brede welvaart in Nederland: nationaal en regionaal. Retrieved from: https://www.uu.nl/sites/default/files/brede_welvaartsindicator_2017_.pdf
- Bavel, B. van, Hardeman, S. and Rijpma, A. (2019). Vervolgstappen voor integrale welvaartsmeting, *ESB 104*, 22-25. Retrieved from: <https://esb.nu/esb/20051494/dossier-meten-van-welvaart>
- Boarini, R. (2011). Measuring Well-being and Progress: the OECD Better Life Initiative, *The Statistics Newsletter - OECD 52*, 3-4. Retrieved from: <https://www.oecd.org/sdd/48387381.pdf>
- Boumans, M. (2001). Fisher's Instrumental Approach to Index Numbers, *History of Political Economy 33*, 313-344. Retrieved from: <https://muse.jhu.edu/article/13328/summary>
- Boumans, M. (2019). De lange geschiedenis van het meten van welvaart, *Economische Statistische Berichten 104*, 53-56. Retrieved from: <https://esb.nu/esb/20051494/dossier-meten-van-welvaart>
- Boumans, M. and Davis, J. (2016). Introduction. In M. Boumans and J. Davis (Eds.), *Economic Methodology: Understanding Economics as a Science: Second Edition*, 1-8. London: Red Globe Press.
- Braithwaite, R. and Roberts, M. (2021). Are Discount Rates Too High? Population Health and Intergenerational Equity, *Medical Decision Making 41*(2), 245-249. Retrieved from: <https://journals.sagepub.com/doi/pdf/10.1177/0272989X20979816>
- Bree, T. van, Jonge, J. de (2022). 'Brede Welvaart' vraagt om een nieuwe definitie van verdienvermogen, *ESB*. Retrieved from: <https://esb-nu.proxy.library.uu.nl/blog/20070294/brede-welvaart-vraagt-om-een-nieuwe-definitie-van-verdienvermogen>
- Cartwright, N. and Runhardt, R. (2014). Measurement. In N. Cartwright & E. Montuschi (Eds.), *Philosophy of social science: a new introduction*, 265-287. Oxford: Oxford University Press.
- CBS. (2018a). *Monitor Brede Welvaart 2018*. Retrieved from: <https://www.cbs.nl/nl-nl/publicatie/2018/20/monitor-brede-welvaart-2018>

- CBS. (2018b). *Monitor Brede Welvaart 2018: Een toelichting*. Retrieved from: <https://www.cbs.nl/nl-nl/publicatie/2018/20/monitor-brede-welvaart-2018>
- CBS. (2019). *Monitor Brede Welvaart & Sustainable Development Goals 2019*. Retrieved from: <https://www.cbs.nl/nl-nl/publicatie/2019/20/monitor-brede-welvaart-sdg-s-2019>
- CBS. (2021). *Monitor Brede Welvaart & Sustainable Development Goals 2021*. Retrieved from: <https://www.cbs.nl/nl-nl/publicatie/2021/20/monitor-brede-welvaart-de-sdg-s-2021>
- CBS. (2022). *Monitor Brede Welvaart & Sustainable Development Goals 2022*. Retrieved from: <https://www.cbs.nl/nl-nl/publicatie/2022/20/monitor-brede-welvaart>
- CBS. (2022b). *Natuurlijk kapitaal en brede welvaart in Nederland: Leefomgeving in verandering*. Retrieved from: <https://www.cbs.nl/nl-nl/longread/rapportages/2022/natuurlijk-kapitaal-en-brede-welvaart-in-nederland>
- Costanza, R., Daly, L., Fioramonti, L., Giovannini, E., Kubiszewski, I., Fogh Mortensen, L., Pickett, K., Vala Ragnarsdottir, K., De Vogli, R. and Wilkinson, R. (2016). Modelling and measuring sustainable wellbeing in connection with the UN Sustainable Development Goals, *Ecological Economics* 130, 350-355.
- Coyle, D. (2014). *GDP: A brief but affectionate history*. Princeton: Princeton University Press.
- Dang, H. and Serajuddin, U. (2020). Tracking the sustainable development goals: Emerging measurement challenges and further reflections, *World Development* 127, 1-20. Retrieved from: <https://www.sciencedirect.com/science/article/pii/S0305750X19301445?via%3Dihub>
- Dorlo, C. (2018, May 16). Het gaat goed, maar met de ene groep wel beter dan met de andere, *NOS*. Retrieved from: <https://nos.nl/artikel/2232088-het-gaat-goed-maar-met-de-ene-groep-wel-beter-dan-met-de-andere>
- Drupp, M., Freeman, M., Groom, B. and Nesje, F. (2018). Discounting Disentangled, *American Economic Journal: Economic Policy* 10(4), 109-134. Retrieved from: <https://www.aeaweb.org/articles?id=10.1257/pol.20160240>
- Edwards, D. (1998). Types of Case Study Work: A Conceptual Framework for Case-based Research, *Journal of Humanistic Psychology* 38(3), 36-70.
- Grashoff, H., Harbers, M., Nijboer, H., Merkies, A., Heerma, P. and Koolmees, W. (2016, April 20). Parlementair onderzoek Breed welvaartsbegrip [Letter of Government]. Retrieved from: <https://www.tweedekamer.nl/sites/default/files/atoms/files/34298-3.pdf>
- Hamilton, K. (2016). Measuring Sustainability in the UN System of Environmental-Economic Accounting, *Environmental Resource Economics* 64, 25-36. Retrieved from: https://ideas.repec.org/a/kap/enreec/v64y2016i1d10.1007_s10640-015-9924-y.html

- Hardeman, S. Aalders, R. and Raspe, O. (2020). De eerste haarscheurtjes in brede welvaart worden zichtbaar, *Rabobank/RaboResearch*. Retrieved from: <https://economie.rabobank.com/publicaties/2020/juli/de-eerste-haarscheurtjes-in-brede-welvaart-woorden-zichtbaar/>
- Heath, J. (2013). Climate Policy: Justifying a Positive Social Time Preference. Retrieved from https://www.academia.edu/3106885/Climate_Policy_Justifying_a_Positive_Social_Time_Preference
- Heath, J. (2016). Caring about climate change implies caring about economic growth. Retrieved from: https://www.academia.edu/30297311/Caring_about_climate_change_implies_caring_about_economic_growth
- Hindriks, J. and Myles, G. (2013). *Intermediate Public Economics: second edition*. Cambridge: The MIT Press.
- Hoekstra, R. (2021). Integreer welvaart in beleidscyclus, *ESB 106*, 46-51. Retrieved from: <https://esb.nu/esb/20065896/integreer-welvaart-in-de-beleidscyclus>
- Iltis, A. and Matthews, K. (2017). NTD policy priorities: Science, values and agenda setting, *PLoS Neglected Tropical Diseases 11*(5), 1-4. Retrieved from: <https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0005431>
- Jacobs, B. (2021). Het beleidsdebat over welvaart mist theoretische fundering, *ESB 106*, 35-39. Retrieved from: <https://esb.nu/esb/20066004/het-beleidsdebat-over-welvaart-mist-theoretische-fundering>
- Kjell, O. (2011). Sustainable Well-Being: A Potential Synergy Between Sustainability and Well-Being Research, *Review of General Psychology 15*(3), 255-266.
- Koopmans, C. (2022). Maatschappelijke kosten-batenanalyse en monitoring van brede welvaart: wat kunnen ze van elkaar leren? *TPEdigitaal 16*(2), 61-71.
- Kordestani, A., Peighambari, K. and Foster, T. (2015). Emerging trends in sustainability research: a look back as we begin to look forward, *International Journal of Environment and Sustainable Development 14*(2), 154-169. Retrieved from: https://www.researchgate.net/publication/275220755_Emerging_trends_in_sustainability_research_a_look_back_as_we_begin_to_look_forward
- Loh, J., Green, R., Ricketts, T., Lamoreux, J., Jenkins, M., Kapos, V. and Randers, J. (2005). The Living Planet Index: using species population time series to track trends in biodiversity, *Philosophical Transactions of the Royal Society B 360*, 289-295. Retrieved from: <https://royalsocietypublishing.org.proxy.library.uu.nl/doi/10.1098/rstb.2004.1584>
- Macfeely, S. (2020). Measuring the Sustainable Development Goal Indicators: An Unprecedented Statistical Challenge, *Journal of Official Statistics 36*(2), 361-378.

- Mair, S., Jones, A., Ward, J., Christie, I., Druckman, A. and Lyon, F. (2017). A Critical Review of the Role of Indicators in Implementing the Sustainable Development Goals, In Leal Filho, W. (eds.), *Handbook of Sustainability Science and Research*, 41-56. Cham: Springer. Retrieved from: https://link.springer.com/chapter/10.1007/978-3-319-63007-6_3
- Meadows, D. H., Meadows, D. L., Randers, J. and Behrens III, W. (1972). *The Limits to Growth: A report for the Club of Rome's project on the Predicament of Mankind*. New York: Universe Books.
- Merry, S. (2011). Measuring the World: Indicators, Human Rights, and Global Governance, *Current Anthropology* 52 (S3), S83-S95. Retrieved from: <http://www.jstor.org/stable/10.1086/657241>
- Michael, B. (2018). What does Brunei teach us about using Human Development Index rankings as a policy tool? *Development Policy Review* 36, O414-O431. Retrieved from: <https://onlinelibrary-wiley-com.proxy.library.uu.nl/doi/pdf/10.1111/dpr.12341>
- Morgan, M. (2009). Seeking Parts, Looking for Wholes, *History of Observation in Economics Working Paper Series*. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1496882
- NOS. (2017, October 27). Noord-Drenthe gezonder, veiliger en gelukkiger dan Randstad, NOS. <https://nos.nl/artikel/2199954-noord-drenthe-gezonder-veiliger-en-gelukkiger-dan-randstad>
- Nowakowska, M. (1977). Methodological Problems of Measurement of Fuzzy Concepts in Social Sciences, *Behavioral Science* 22(2), 107-115.
- OECD. (2022, June 10). *What's the Better Life Index? Frequently Asked Questions*, OECD Better Life Index. <https://www.oecdbetterlifeindex.org/about/better-life-initiative/>
- Porter, T. (1995). *Trust in Numbers*, Princeton: Princeton University Press.
- Provinciale Staten van Noord-Brabant. (2017). PS 45/17 Notulen van de vergadering van Provinciale Staten van Noord-Brabant op 21 april 2017. Retrieved from: <https://www.brabant.nl/bestuur/provinciale-staten/statenstukken/besluiten-ps?p=32>
- Ravallion, M. (2012). Mashup Indices of Development, *The World Bank Research Observer* 27(1), 1-32. Retrieved from: <https://www.jstor.org/stable/23262819>
- Reinsdorf, M. (2007). Axiomatic Price Index Theory. In M. Boumans (ed.), *Measurement In Economics: A Handbook*, 153-188. Amsterdam: Elsevier Academic Press.
- Rijpma, A., Moatsos, M., Badir, M. and Stegeman, H. (2017). Netherlands beyond GDP: A Wellbeing Index, *Munich Personal RePEc Archive No 78634*. Retrieved from: <https://mpa.ub.uni-muenchen.de/78934>

- Roemer, J. (2011). The ethics of intertemporal distribution in a warming planet, *Environmental and Resource Economics* 48, 363-390. Retrieved from: <https://link.springer.com/article/10.1007/s10640-010-9414-1>
- RTL. (2017, October 27). 'In de Randstad woon je ongelukkig, onveilig en ongezond', *RTL Nieuws*. Retrieved from: <https://www.rtlnieuws.nl/nederland/artikel/3710716/de-randstad-woon-je-ongelukkig-onveilig-en-ongezond>
- Sen, A. (2000). *Development as Freedom*. New York: Alfred A Knopf.
- Smit, P. (2022, June 12). Op 'brede welvaart' - de meter voor kwaliteit van leven - scoort de stad Zwolle hoog. Maar voor hoe lang nog? *De Volkskrant*. Retrieved from: <https://www.volkskrant.nl/nieuws-achtergrond/op-brede-welvaart-de-meter-voor-kwaliteit-van-leven-scoort-de-stad-zwolle-hoog-maar-voor-hoe-lang-nog~b2ca4be2/>
- Starr, M. (2014). Qualitative and Mixed-Methods Research in Economics: Surprising Growth, Promising Future, *Journal of Economic Surveys* 28(2), 238-264. Retrieved from: <https://onlinelibrary-wiley-com.proxy.library.uu.nl/doi/pdf/10.1111/joes.12004>
- Stiglitz, J., Sen, A. and Fitoussi, J. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Retrieved from: <https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf>
- United Nations. (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development*. Retrieved from: <https://sdgs.un.org/2030agenda>
- United Nations (2017). Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development. Retrieved from: https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%202019%20refinement_Eng.pdf
- United Nations (2022a). Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development. Retrieved from: <https://unstats.un.org/sdgs/indicators/indicators-list/>
- United Nations. (2022b, June 12). *SDG Indicators*. <https://unstats.un.org/sdgs/indicators/indicators-list/>
- United Nations. (2022c, June 12). *The Sustainable Development Agenda*. <https://www.un.org/sustainabledevelopment/development-agenda/>
- UNDP. (1990). *Human Development Report 1990*. Oxford: Oxford University Press.
- UNSD. (2015). United Nations Fundamental Principles of Official Statistics: Implementation Guidelines. Retrieved from: https://unstats.un.org/unsd/dnss/gp/Implementation_Guidelines_FINAL_without_edit.pdf

- UNSD. (2022, June 13). IAEG-SDGs: Tier Classification for Global SDG Indicators. <https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/>
- UNECE/OECD/Eurostat. (2014). Conference of European Statisticians Recommendations on Measuring Sustainable Development. Retrieved from: https://unece.org/DAM/stats/publications/2013/CES_SD_web.pdf
- Vollebergh, H. (2021). Beleidsdebat brede welvaart prima theoretisch gefundeerd, *ESB 107*, 164-167. Retrieved from: <https://esb.nu/esb/20069596/beleidsdebat-brede-welvaart-prima-theoretisch-gefundeerd>
- Vos, R. (2007). Defining sustainability: a conceptual orientation, *Journal of Chemical Technology and Biotechnology* 82, 334-339.
- Vucetich, J. and Nelson, M. (2010). Sustainability: Virtuous or Vulgar? *Thinking of Biology* 60(7), 539-544.
- Wagle, U. (2005). Multidimensional Poverty Measurement with Economic Well-being, Capability, and Social Inclusion: A Case from Kathmandu, Nepal, *Journal of Human Development* 6(3), 301-328. Retrieved from: <https://www.tandfonline.com/doi/full/10.1080/14649880500287621>
- Wagner, G., Anthoff, D., Cropper, M., Dietz, S., Gillingham, K., Groom, B., Kelleher, P., Moore, F., and Stock, J. (2021). Eight priorities for calculating the social cost of carbon, *Nature* 590, 548-550. Retrieved from: <https://www.nature.com/articles/d41586-021-00441-0>
- World Commission on Environment and Development. (1987). *Our common future*. Oxford: Oxford University Press.
- Zeller, R. and Carmines, E. (1980). *Measurement in the social sciences*. Cambridge: Cambridge University Press.