Aus dem
Deutschen Krebsforschungszentrum (DKFZ) Heidelberg
(Wissenschaftlicher Vorstand: Prof. Dr. med. Michael Baumann)

Exploring Solutions to Conceptual and Measurement Problems in the Capability Approach: the Development of the WeRFree Instrument

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)
der
Medizinischen Fakultät Mannheim
der Ruprecht-Karls-Universität
zu
Heidelberg

vorgelegt von
Jasper Tobias Ubels

aus
Den Helder
2023

Dekan: Prof. Dr. med. Sergij Goerdt
Referent: Prof. Dr. med. Michael Schlander

TABLE OF CONTENTS

ABBREVIATIONS

15D	Fifteen Dimensional measure of health-related quality of life
ASCOT	Adult Social Care Outcomes Toolkit
ANOVA	ANalysis Of Variance
APHAB	Abbreviated Profile of Hearing Aid Benefit
AQoL(-4D)(-8D)	Assessment of Quality of Life instrument (- four dimensions) (-eight dimensions)
CALY	Capability-Adjusted Life-Year
CFI	Comparative Fit Index
CGM	Continuous Glucose Monitoring
CoCan-ELSI	The Impact of the Corona Crisis on Cancer Patients: medical, psychosocial and economic consequences and their ethical implications
COREQ	Consolidated Criteria for Reporting Qualitative Research
COVID-19	COronaVIrus Disease 2019
df	degrees of freedom
DKFZ	German Cancer Research Center
DWLS	Diagonally Weighted Least Squares
EQ-5D-5L	EuroQol five-dimensional instrument with five response options
FIML	Full Information Maximum Likelihood
HRQoL	Health-Related Quality of Life
HTA	Health Technology Assessment
HUI-3	Health Utilities Index Mark 3
ICECAP-A	ICEpop CAPability measure for Adults
ICECAP-O	ICEpop CAPability measure for Older people
ICECAP-SCM	ICEpop CAPability Supportive Care Measure
ICER	Incremental Cost-Effectiveness Ratio
KID	KrebsInformationsDienst (Cancer information center)
MIC	Multi Instrument Comparison
MLR	Robust Maximum Likelihood
NCT	National Center for Tumor Diseases Heidelberg
NIPT	NonInvasive Prenatal Testing
ONS	Integrated household survey from the Office of National Statistics
OCAP-18	Instrument developed to measure Nussbaum's list of capabilities
OxCAP-MH	Oxford CAPabilities questionnaire-Mental Health
QoL	Quality of Life
PWI	Personal Wellbeing Index
QALYs	Quality Adjusted Life Years
QWB-SA	Quality of Well Being Self-Administered scale
RMSEA	Root Mean Squared Error of Approximation
SCRQoL	Social Care Related Quality of Life
SF-36 V2	36-Item Short Form Health Survey Version 2
SF-6D	Short-Form Six-Dimension
SG	Standard Gamble
SRMR	Standardized Root Mean Residual
SWLS	Satisfaction With Life Survey
TLI	Tucker Lewis Index
TTO	Time Trade-Off
WeRFree	Wellbeing Related option Freedom
WCI	Women´s Capability Index

1 INTRODUCTION

1.1 Structure of this dissertation

The general subject of this dissertation is how to apply the capability approach in health economics. Specifically, this dissertation concerns the measurement of wellbeing with questionnaires that are completed by individuals evaluating their own lives. These questionnaires are also called “self-report instruments” or, in this dissertation, simply “instruments”. The dissertation is structured as follows to address this subject.

In Chapter 1, a selection of wellbeing theories that are necessary to understand the foundations of how the effects of interventions are assessed in health economics will be introduced and shortly discussed. Chapter 2 is a review of how researchers have operationalized the capability approach for measuring wellbeing in the field of health research. In Chapter 3, an alternative and potentially improved framework for operationalizing concepts from the capability approach for wellbeing measurement is developed. In Chapter 4, this framework is applied in a psychometric analysis to establish whether the theoretical concepts can be translated into measurable constructs. Furthermore, an instrument is developed in this chapter. In Chapter 5, I illustrate how measurement invariance testing can be used to study whether advantaged and disadvantaged groups of individuals interpret and respond to capability instruments in a similar way. This dissertation ends with a discussion in Chapter 6, in which I address the overall limitations of the work presented in this dissertation. Furthermore, I provide an overview of the research that is needed before the instrument that is developed in this dissertation can be used in practice.

1.2 Background

At its very core, this dissertation asks the question what kind of information is needed to determine whether a health intervention is valuable. The question whether something is valuable is closely linked to the question whether something is good or bad (Schroeder, 2021). To illustrate this point, a typical expression might be “happiness is good”, which indicates that happiness has value. Stipulating whether something is good or bad (i.e., happiness or sadness) is a value judgement (Schroeder, 2021). Value judgements about the type of information that is relevant for

the assessment of value are therefore moral claims about what ought to be good or bad (Scanlon, 1991; Schroeder, 2021).

Such value judgements are also being made to determine whether a health intervention is valuable (Hofmann, 2005). For example, to study the effectiveness of a health intervention, an outcome has to be defined and measured (Hofmann, 2005). Usually, these outcomes are health-related, which means that the "goodness" of an intervention is based on a value judgement about what counts as "health". Such value judgements about health are closely associated with moral claims about overall wellbeing (Schroeder, 2021). To illustrate, when something is argued to be healthy, for example low blood pressure, it is implicitly or explicitly implied that this is good for an individual's wellbeing. As such, evaluations of health interventions involve implicit or explicit claims about the wellbeing of individuals.

A particular practice in which such value assessments are conducted is health economic evaluation. A health economic evaluation has the aim to inform decision-makers about how resources should be allocated to reach certain outcomes (Drummond et al., 2015a). Such an evaluation is conducted by comparing the benefits and costs of investing in different health interventions. Benefits can be measured in terms of a variety of outcomes. These outcomes can be immediate reflections of health, such as a reduction in morbidity or mortality, but also reflect broader aspects, such as health-related quality of life (HRQoL).

Over the last decades, health economic evaluations have been increasing in influence as well as in number (Drummond et al., 2015a). According to Drummond et al. (2015a), this increase was driven by two factors. First, healthcare budgets are increasingly under pressure. Second, several jurisdictions have developed processes to evaluate health interventions that use health economic evaluations in decision making.

These processes are also called "health technology assessment" (HTA). Technology assessment can be seen as a form of policy research that aims to understand the consequences of applying technologies (Banta, 2009). HTA has been defined as the practice of bridging the gap between evidence and decision-making in the field of healthcare (Banta, 2003). HTA is conducted by HTA agencies, which stipulate what

kind of information, in the form of evidence, should be used for value assessments of health technologies.

Studies have shown that for the same health technology, different agencies come to different conclusions in their value assessments, which affects the access of patients to new health technologies (Schaefer et al., 2021; Schaefer and Schlander, 2019). In part, this could be a consequence of the agencies' choices regarding the type of evidence that is accepted to assess the effects of health technologies and inform value assessments (Schaefer et al., 2021; Schaefer and Schlander, 2019). As mentioned, a decision regarding what type of information should be used for value assessment is essentially a value judgement regarding what elements are good for an individual's wellbeing. Thus, I will first introduce how wellbeing is understood, measured, and evaluated in health economics.

1.3 Wellbeing in health economics

Gasper (2010) defines wellbeing as "... an abstraction that is used to refer to the quality of any of many valued aspects of (a) life, or some set thereof, or their totality". A multitude of theories of wellbeing exist. It goes beyond the scope of this dissertation to give a complete review of these theories, but overviews can be found in Gasper (2010) and Phillips (2006). Instead, I will introduce a selection of wellbeing theories that help understand the information (or informational base) that is used in health economics to assess the effect and value of health technologies. The introduction of these wellbeing theories will be loosely structured around Sen's analysis of the concept "utility". Sen distinguishes between three different conceptualizations of utility (Sen, 1985a):

1. Utility as a form of happiness.
2. Utility as a form of desire fulfillment.
3. Utility as reflecting choice.

The introduction of these concepts of utility will be followed by Sen's critique of these theories of wellbeing and the introduction of the capability approach.

1.3.1 Happiness

According to Sen, utility as a form of happiness conceptualizes utility as consisting of different emotional experiences (Sen, 1985a, 1987). Examples of such experiences

are happiness itself, but also excitement or stimulation (Sen, 1985a). According to Sen, these emotional experiences are important for human wellbeing.

Beyond these examples, Sen does not clearly define or differentiate between different types of happiness and why they might be relevant for wellbeing (Clark, 2005). To provide a further background to some of the discussions in this dissertation, I therefore introduce some happiness accounts from philosophy and psychology. First, I will introduce two concepts with a long tradition in western philosophy: hedonia and eudaimonia (Huta and Waterman, 2014). Then, I will introduce the concept “subjective wellbeing” (Dolan and White, 2007).

In hedonic theories, the only type of information that is relevant for the assessment of wellbeing is the subjective experience of an individual, with a particular focus on pleasure (Huta and Waterman, 2014; Moore, 2004; Ryan and Deci, 2001). Depending on the specific hedonic theory, also further subjective experiences could be relevant for the evaluation of wellbeing, such as enjoyment, happiness, life satisfaction, low levels of distress, relaxation, and comfort (Huta and Waterman, 2014). In the context of the concept of utility, strong parallels can be drawn between the traditional understanding of utility and hedonism (Kymlicka, 2002). Bentham defined utility as consisting of pleasure and the absence of pain (Bentham, 1781). Authors in traditional utilitarianism have argued that policies should be evaluated according to their effect on promoting pleasure or reducing pain (see for example Mill (1887)).

The principle that underlies the eudaimonic take on wellbeing and happiness stipulates that people should live in a certain way to “flourish” (Huta and Waterman, 2014; Ryan and Deci, 2001). People who live according to this prescribed way of life may be considered to have achieved eudaimonia, which is a word that could be translated as “happiness” or “wellbeing”. Historically, the *experience* of some kind of pleasure or comfort while living a flourishing life was considered to be a bonus, but not a necessity for someone to be considered to live a eudaimonic life (Ryan and Deci, 2001). Thus, historic understandings of eudaimonia had little to do with the evaluation of subjective experiences. However, in contemporary work, authors have argued that some subjective experiences can be associated with flourishing or living a virtuous life. In these contemporary conceptualizations of eudaimonia, commonly shared subjective experiences that are used as an informational base to assess happiness include, but

are not limited to, the development of potential, having a purpose, and identity (Huta and Waterman, 2014). Consequently, the contemporary understanding of eudaimonia can be applied as a self-reported outcome in research and does not only function as a normative prescription of the components that are key to a virtuous life (Huta and Waterman, 2014). A questionnaire that is commonly used by researchers to measure eudaimonic wellbeing is Ryff (1989) scale of psychological wellbeing.

Another understanding of happiness besides eudaimonia and hedonia is that of subjective wellbeing. Diener (1984) defined subjective wellbeing as the experience of individuals that their life is going well. This includes both negative and positive experiences. The assessment of these experiences is usually an integrated, global judgement, which goes beyond the assessment of wellbeing in specific domains. Subjective wellbeing is argued to consist of three components: positive affect, negative affect, and life satisfaction (Busseri and Sadava, 2011; Diener, 1984). Positive and negative affect can be seen as the emotional components of wellbeing. As a component of wellbeing, positive affect represents those emotions that can be considered “positive”, such as happiness or joy (Diener, 1984). Negative affect, conversely, represents emotions that can be considered “negative”, such as sadness or anger (Diener, 1984). Life satisfaction is understood as a cognitive evaluation of an individual over his or her own life. As a cognitive evaluation, life satisfaction might influence positive and negative affect but is itself not a direct measure of affect (Diener, 1984). The exact way in which positive affect, negative affect, and life satisfaction are related to each other (for example, whether these constructs should be understood as a causal system or form a hierarchical structure) is still under debate (Busseri and Sadava, 2011).

Several differences can be identified when comparing the different theories of happiness and subjective wellbeing. One major difference between the hedonia, eudaimonia, and the happiness understandings of utility theories on the one hand and subjective wellbeing on the other is that the former theories are essentially based on moral philosophy, while the latter theory is based on empirical research (Diener et al., 2009). Applications of hedonia, eudaimonia, and the happiness understanding of utility thus reflect moral principles from their respective research traditions about what kinds of information *should* be included in happiness assessment. In contrast, subjective wellbeing research is grounded in empirical psychology and aims to *describe and*

understand the experience of happiness (Diener et al., 2009). One further difference is that the happiness understanding of utility not only normatively describes what kind of information should be used for wellbeing assessment (i.e., a specific understanding of happiness), but also how this information should be used to inform policymaking (Gasper, 2010).

Aside from these differences, it should also be noted that these theories of happiness have much in common. From a theoretical perspective, the contemporary understanding of these theories generally shares the assumption that wellbeing can (or even should) be assessed on the basis of the subjective experiences of individuals (Gasper, 2010). Furthermore, it should also be noted that the differences between these theories of happiness are less clear in practice. Authors have conceptualized and consequently operationalized these theories in a variety of different ways (Huta and Waterman, 2014). This has led to measures that are based on different theories of happiness showing similarities in terms of their content (Huta and Waterman, 2014).

To conclude, theories and empirical findings regarding the nature of happiness and subjective wellbeing provide a deep understanding of which experiences might be relevant for the assessment of overall wellbeing. Insights from these fields of research will facilitate the interpretation of some of the results of this dissertation. However, for now I will continue with the introduction of alternative conceptualizations of utility.

1.3.2 Desire and choice

In the “desire” understanding of utility, desire fulfillment is seen as something good. The value of an object is based on information about the desirability of that good to an individual (Sen, 1992b). Utility is achieved when a desired state is, in Sen’s words, objectively realized. To make interpersonal comparisons, further information is however needed to assess the value of a good. One way is to measure and compare the intensity of the desire of different individuals for different goods (Sen, 1992b). This intensity functions as a measure of the strength of a desire for a specific object. Based on this information, it is theoretically possible to make interpersonal comparisons of wellbeing.

In the choice understanding of utility, utility is understood as a representation of the choice behavior of an individual (Sen, 1985a). The value of a good is determined with

information about an individual's choice behavior. The underlying assumption is that the observable choices of people are guided by their pursuit of personal wellbeing (Sen, 1985a). By studying the choices that individuals make from a set of alternatives, it is possible to create an "ordinal scale" of those alternatives, with alternatives higher in the ranking yielding more utility, and thus having more value, than alternatives lower in the ranking (Sen, 1985a). However, the utilities of the goods in this set cannot be expressed numerically if only information is available about the relative ranking of the alternatives in a set. In a more stringent framework with additional assumptions, it is possible to create a "cardinal scale" of those alternatives (Sen, 1985a). On a cardinal scale, the utilities associated with the alternatives have numerical values, which represent the value of those alternatives.

Sen observed that the use of choice information is particularly popular in the field of economics (Sen, 1985a, 1987). This is also the case in the field of health economics. Utility understood as "choice" forms the basis for normative frameworks that guide decision-making in healthcare investment. In conventional health economics, it is possible to broadly label two frameworks of assessing value: welfarism and extra-welfarism.

1.3.2.1 Welfarism in health economics

The normative foundations of the economic evaluation of health technologies can be found in welfare economics (Coast et al., 2008c). According to Hurley (2000), welfare economics is based on four key principles:

1. The maximization of utility, which represents the idea that individuals can rationally rank options from a given set according to their preferences. As such, it shows parallels with what Sen would describe as the "choice" understanding of utility. From observing this ranking, one can infer the relative value of those options to that individual. This ranking has to be done consistently because otherwise, the observed ranking becomes meaningless.
2. Individual sovereignty, which expresses that individuals themselves are the experts in evaluating their own welfare.
3. Consequentialism, which asserts that, in the context of health economics, the effect of health technologies can only be evaluated in terms of their outcomes. The process itself that led to an outcome is, according to this principle, irrelevant.

4. Welfarism, which represents the idea that the desirability of any situation can only be judged in terms of the impact on individual utilities in that situation. Any other type of information is excluded from this judgement.

A distinction can be made between classical and neo-classical welfarism (Brouwer et al., 2008; Hurley, 2000). To aggregate utilities in classical welfarism, it is assumed that interpersonal utility values are comparable and that it is possible to give a specific utility value to different goods from a set (Brouwer et al., 2008; Hurley, 2000). In neo-classical welfarist theory, this assumption is however dropped, which means that interpersonal comparisons of utility cannot be made. This means, that single goods do not receive a specific utility value, since the utility value of that good to an individual is incomparable across individuals (Brouwer et al., 2008). Given that single goods do not have a specific utility value, it is also not possible to attribute a numerical utility value based on the set of goods available to an individual as a representation of the wellbeing of that individual. From this follows the principle that the welfare of a group cannot be expressed in terms of the aggregate of the obtained utilities from the members of that group. As such, no value judgements can be made about how to distribute resources among different groups in society in terms of aggregate utilities.

Instead, the Pareto principle can be applied to evaluate the distribution of goods in a society in neo-classical welfarism (Hurley, 2000). According to the Pareto principle, the distribution of goods in a society is Pareto optimal when no change in the distribution of goods can lead to an individual in society being better off without making another individual in that society worse off (Hurley, 2000; Schlander, 2005). Of course, this principle is very strict. Within the context of health care, any kind of policy that redistributes resources from one individual to the next (such as taxes or mandatory health insurance memberships) violates the Pareto principle (Hurley, 2000; Schlander, 2005). For this reason, the “potential Pareto improvement” principle is used in practice (Hurley, 2000; Schlander, 2005). According to this principle, the welfare of a society can still improve when an individual that is harmed by, for instance, a policy that benefits others could be compensated in such a way that the individual at least has the potential to reach an equal level of welfare as before the implementation of that policy (Hurley, 2000; Schlander, 2005).

1.3.2.2 Applications of welfarism in health economics

In applications of welfarism, the information that is used to assess the value of a good is the willingness to pay (WTP) of an individual for that good (Hurley, 2000). This monetary amount is an expression of the utility that a good yields to an individual. Some goods can however not directly be bought on the market. An example of such a good from health economics is a unit of “health” (Hurley, 2000). In this context, a unit of health is an abstraction of the effect of various health-related services that can be consumed, such as medical treatments.

Given that a market to value an abstract good such as health does not exist, researchers developed several valuation methodologies to come to WTP estimations for health (Breidert et al., 2006). A general distinction can be made between valuation methodologies based on revealed preferences and valuation methodologies based on stated preferences. Observations about market behavior are the source of WTP information for revealed preference approaches. By using either market data or experiments the WTP for abstract goods such as health can be established (Breidert et al., 2006). To do so, abstract goods such as health need to be operationalized as a tangible outcome (for example mortality).

Survey data are the basis for stated preference methods for eliciting WTP. These surveys can be constructed in such a way that they directly inquire about the WTP of individuals for a certain good (Breidert et al., 2006). Again, health will need to be operationalized, for example utilizing a description of a health state. A health state description reflects various components that can be associated with being healthy, such as not being in pain, being mobile, and not being depressed (Hurley, 2000). Individuals can then be asked how much they would be WTP to not be in pain, be mobile, and not depressed (Hurley, 2000).

Alternatively, surveys can be constructed in such a way that they indirectly estimate a WTP for a good. These surveys can be categorized as either conjoint analysis or discrete choice analysis (Breidert et al., 2006). Competing sets of goods are described, in both these types of surveys, with additional information about the price for the full set. Individuals are asked to indicate which set of goods they prefer. In the context of health, these descriptions of sets could be a description of two health states. Health state one describes an imperfect level of health, representing some kind of disease

with no extra costs involved. Health state two could describe an “ideal” health state, with the additional information that the achievement of this health state costs a certain amount of money. Individuals are then asked to choose which health state they prefer. Information from this choice behavior can be used to estimate the WTP for a certain improvement in health. These types of indirect surveys are typically used to estimate WTP in health economics (Hurley, 2000).

In health economic evaluations, information about the WTP for a unit of health is used to inform cost-benefit analysis (Hurley, 2000). In cost-benefit analysis, the per-patient costs of the implementation of a health technology are subtracted from the individual’s WTP for the health improvement that is a result of that intervention. The reimbursement of a health technology is considered desirable when, over a group of people, the aggregated monetary benefits are higher than the aggregated monetary costs (Hurley, 2000). The implementation of such a health technology would be a potential Pareto improvement: on a societal level, the costs that are created can be compensated by the (potential) benefits that a health technology yields (Hurley, 2000). The principle aim of cost-benefit analysis is to maximize aggregate utility (expressed in monetary units) in society.

However, welfarism has been critiqued on various grounds. Hurley (2000) lists three key critiques that have been raised in literature that are particularly relevant for health economics.

1. The first critique is that some assumptions of welfarism do not seem to hold. For example, some authors have argued that individuals do not have the expertise to evaluate their own welfare in the context of health (Culyer, 1989; Hurley, 2000; Sen, 2002), which thus invalidates the principle that only information from individuals themselves is relevant for the assessment of their wellbeing.
2. The second critique is that WTP itself is not an accurate representation of the value of a health technology, given that health is unique in the sense that it affects an individual´s very existence (Hurley, 2000). As such, the WTP for a unit of health might be limited by the financial resources that individuals have at their disposal, or, in other words, WTP might represent an individual’s ability to pay (Brouwer et al., 2008; Coast et al., 2008c). Consequently, the monetary value that individuals are willing (and able) to pay might not reflect the actual utility value of those

interventions to those individuals, since individuals would pay more if they could (Coast et al., 2008c).

3. The third and last critique is that authors have argued that not only utility, but also health itself is a variable that is of direct importance to society (Coast et al., 2008c). As such, the relevant informational base to inform economic analysis in the field of health should not be utility, but health (Culyer, 1989; Hurley, 2000).

These critiques urged authors to develop alternative evaluative frameworks, leading to the emergence of extra-welfarism.

1.3.2.3 Extra-welfarism in health economics

The notion that health itself is of direct importance to society is also reflected in earlier writings within the health economic field (Coast et al., 2008c). This led to the development of the "decision-makers approach" (Coast et al., 2008c). In this approach, those objectives that are considered to be important by the responsible decision-maker are also the societal objectives (Coast et al., 2008c). In the context of healthcare, this meant that economists and decision analysts argued that what mattered for decision-makers is the maximization of health, which consequently meant that health itself should be the subject and outcome of choice for value assessments instead of utility (Coast et al., 2008c).

Initially, the decision-makers approach was not based on a theoretical framework, which was a point of critique by welfarists (Brouwer and Koopmanschap, 2000; Coast et al., 2008c). However, further developments of the approach led to the extra-welfarism framework (Brouwer and Koopmanschap, 2000; Coast et al., 2008c). But, what makes extra-welfarism "extra"?

According to Brouwer et al. (2008), four differences between welfarism and extra-welfarism can be identified.

1. In extra-welfarism, the informational base for (health) economic evaluations is extended, which means that outcomes other than utility can be used for evaluative purposes. Examples of such outcomes are health, stigmatization by society, or the quality of friendships (Culyer, 1989). In the context of health economic analysis, extra-welfarists have argued that health should be the primary outcome because

the main goal of the health care system is to maximize its health output (Culyer, 1989; Hurley, 2000).

2. Not only information from the individuals affected by a policy decision can be used to construct an informational base for evaluation. Instead, other sources of evidence can be used as well. In the context of the health sector, this means that information derived from the general population can be used to inform decisions that affect patients (Brouwer et al., 2008).
3. In extra-welfarism (societal) outcomes can be valued with methods that are not necessarily preference-based (Brouwer et al., 2008). Instead, the valuation of (societal) outcomes can be based on other sources of information, such as the productivity of an individual, or on ethical principles besides the maximization of a certain outcome (Brouwer et al., 2008), such as the "fair-innings" principle by Williams (1997).
4. In the extra-welfarist framework, it is possible to interpersonally compare the wellbeing of individuals with a variety of different dimensions, as opposed to the welfarist approach that focuses only on utility. This is a consequence of the extra-welfarist's flexibility regarding the types of information that are valid for constructing an informational base for decision-making. For example, within extra-welfarism, it is possible to say that a person affected by a disease has a lower level of wellbeing than a healthy person, even though we do not have information about these individuals' respective utilities. This makes it theoretically possible to directly maximize certain outcomes in a population, such as health (Brouwer et al., 2008).

1.3.2.4 Applications of extra-welfarism in health economics

In health economics, extra-welfarism forms the theoretical foundation of cost-utility analysis (Hurley, 2000; Schlander, 2005). In cost-utility analysis, Quality Adjusted Life Years (QALYs) are the informational base that reflects the effect of health technologies on the wellbeing of patients (Torrance, 1986). QALYs are composed of life years with a quality of life (QoL) adjustment (Schlander, 2005). To calculate the QALYs of an individual, the time that an individual lives in a certain health state is multiplied by a preference-adjusted QoL score of that health state (Torrance, 1986).

Torrance (1986) distinguishes three different methods to develop a description of health states that are used for the aforementioned preference adjustments. In the first method, researchers ask patients to value their own state of health. In this case, patients could be asked to write down their experiences to facilitate the interpretation

of the results of the preference elicitation. The second method is to describe health states holistically, which can take multiple forms. One way is to provide a couple of key sentences to describe a certain health state. Alternatively, it is also possible to write a narrative of a couple of paragraphs. These descriptions are then used to elicit preferences from participants who are not affected by the disease described in the health state.

The third, method of developing health state descriptions is by basing them on a health state classification system. The health state classification system is based on the concept that a construct, such as health, can be described with several attributes. These attributes are different elements related to health, such as pain, emotional functioning, or mobility. Per attribute, different levels will need to be defined. For example, for the attribute pain, it is possible to develop the following three levels: (1) no pain, (2) a bit of pain, and (3) a lot of pain. In conventional health economic analysis, health states are usually based on the content of questionnaires called “multi-attribute utility instruments” (Torrance, 1986). Values are then assigned to different health states in such instruments, which represent the QoL of living in a particular health state. These values to different levels within attributes are also called weights (Decancq and Lugo, 2013).

Once a health state has been described, it also has to be valued. In extra-welfarism, this is done by eliciting preferences. A multitude of methods can be used to elicit preferences for health state descriptions, but in extra-welfarism there are three classic methods: the rating scale, the standard gamble (SG), and the time trade-off (TTO) (Torrance, 1986). The rating scale is the most straightforward method: it is a straight line, with a favorable health state on one side and an unfavorable one on the other, which function as “anchors” (Torrance, 1986). Between the ends of this line, it is also possible to describe additional health states which are ordered in terms of their favorability. Individuals are then asked to score a health state on this line, which can either be a description of a hypothetical health state or an actual experienced health state by the individual herself. The point that the individual chooses on this line is a score that can be interpreted as the utility value of that point relative to the health states at both ends of the line (Torrance, 1986).

The TTO is a slightly more complex method of eliciting preferences that involves a choice. In the TTO, two scenarios are presented to individuals (Torrance, 1986). In the first scenario, a health state *i* is described in which the individual has to hypothetically live for a *t* amount of time, after which he or she dies. This health state is worse than an "ideal" health state and might involve a description of an illness. In the second scenario (Torrance, 1986), an ideal health state is described in which the individual has to live in $x < t$ amount of time, which is also followed by death. The timeframe *x* is then varied to find the value where an individual is indifferent between the scenarios and cannot choose which one is more favorable. The preference value for that specific health state can then be expressed as $h_i = x/t$ (Torrance, 1986).

The SG is also a method that involves a choice (Torrance, 1986). Again, two scenarios are presented to an individual. In the first scenario, an imperfect health state is described in which an individual has to hypothetically live for a number of years. In the alternative scenario, a treatment is presented which gives individuals the ability to become healthy (for the same number of years as scenario one). This treatment comes however with a caveat: there is a *p* chance that it kills the individual (Torrance, 1986). Probability *p* is then varied until an individual is indifferent between the two scenarios. When this happens, probability *p* presents the utility value for the health state presented in scenario one (Torrance, 1986).

In extra-welfarism, it is not directly possible to subtract costs and effects, since effects are expressed in terms of QALYs instead of a monetary value. Therefore, the incremental cost-effectiveness ratio (ICER) is used as an expression of value. An ICER describes the relationship between costs and effects. The ICER is calculated as follows (Rudmik and Drummond, 2013; Schlander, 2005):

$$ICER = \frac{C_1 - C_2}{E_1 - E_2}$$

Where C_1 is the cost of health technology one, C_2 is the cost of health technology two, E_1 is the effect of health technology one, and E_2 is the effect of health technology two. In extra-welfarism, effects are expressed in terms of QALYs. The resulting ICER provides is an expression of the marginal value of the health technologies under comparison. The ICER can be compared to a cost-effectiveness threshold to assess

whether a new health technology should be reimbursed (Rudmik and Drummond, 2013).

Theoretically, an evaluation based on the extra-welfarist framework could be very different compared to an evaluation that is based on the welfarist framework. Due to its flexibility, an extra-welfarist framework could explicitly incorporate equity assessment in its evaluations (Brouwer et al., 2008). In practice, however, there are large similarities between welfarism and extra-welfarism (Hurley, 2000). Both frameworks have an almost exclusive focus on maximizing a single aggregate outcome to evaluate the effect of health technologies. This outcome is based on information that is, in theory, linked to the preferences of individuals (Hurley, 2000). Lastly, both share a strong consequentialist approach to the evaluation of health technologies, which leaves little room for concerns related to whether the evaluation process itself is justified (Hurley, 2000). For this reason, welfarism and extra-welfarism are critiqued on similar grounds (Coast et al., 2008c; Hurley, 2000).

To summarize this section, I have introduced different frameworks that have been used to evaluate the wellbeing of individuals. What these frameworks share is that the information that forms the basis of these evaluative judgements is derived from individuals, either through direct reports or through observation of choice behavior (Gasper, 2010). Much trust is thus placed on the quality of information that can be derived from individuals. Furthermore, both the welfarist and extra-welfarist argue for the maximization of wellbeing in a population.

However, it should be noted that authors have questioned the reliability of information derived from individuals to evaluate their wellbeing. Furthermore, the focus on maximizing outcomes has also been questioned. One of the most famous critics of particularly the welfarist framework is Amartya Sen. His critique and the alternative framework that he developed will be the subject of the next section.

1.4 Limitations of self-reported information and adaptation

In the last section, I introduced how Sen identifies three different understandings of utility that can be used as the informational basis for the evaluation of wellbeing: choice, happiness, and desire. According to Sen, each of these three understandings is limited in the context of wellbeing assessment, which consequently means that each

of the three only provides a limited informational basis to assess how well-off individuals are (Sen, 1999a). I will shortly discuss why Sen considers each understanding of utility to be limited.

Regarding the desire understanding of utility, Sen recognizes that desire may be closely linked to value (Sen, 1985a, 1987). Indeed, he notes that something of value might plausibly cause desire. However, Sen argues that value and desire are different concepts. Indeed, he continues to argue that the idea that something is valuable *because* of an individual's desire does not make much sense (Sen, 1985a, 1987). Sen illustrates this point with an example. Imagine an individual who asks why x has a certain value. The answer "because someone desires it" is insufficient to explain the actual value of x. Sen thus concludes that desire can hardly be seen as the basis for value. Further discussions can be found in Sen (1985a) and Sen (1987).

Also, the use of information about individuals' choice behavior has its limitations when it comes to the assessment of wellbeing and the valuation of goods (Sen, 1985a, 1987). From a theoretical perspective, Sen argues that choice behavior and individual benefit are two separate concepts that are confounded but do not share a direct link (Sen, 1987). Sen further argues that the only way that choice and wellbeing could be connected is if choice itself is some kind of reflection of desire (Sen, 1987). Consequently, Sen's critique on the desire understanding of utility can also be applied to the choice understanding of utility (Sen, 1987). One further critique on the choice understanding of utility is that an individual's choice can be motivated on different grounds beside the fulfillment of his or her own wellbeing. Some choices by individuals go directly against their own wellbeing, which means that the assumed link between choice behavior and individual benefit itself is at least questionable (Sen, 1985a). Consequently, information about choice behavior can at best provide a limited amount of information about an individual's wellbeing (Sen, 1985a, 1987). Additionally, empirical studies show that the principles of utility theory (see principles by Hurley (2000) earlier this chapter) do not adequately describe individuals' choice behavior under risk, which particularly affects the interpretation of SG values (Kahneman and Tversky, 1979).

Lastly, there is also the adaptation problem, which limits the usefulness of utility as an informational input in each of its three conceptualizations but has particularly strong

implications for the happiness understanding of utility (Sen, 1985a, 1987). One key problem with the sole use of utility as an informational base is that people adapt their preferences to disadvantage. According to Sen, someone who is disadvantaged might still report high levels of happiness or desire fulfillment, even though this person would not be considered to have a high level of wellbeing by an outside observer (Sen, 1985a, 1987). Certain forms of adapted preferences can even be seen as irrational. These irrationally adapted preferences have also been called adaptive preferences by Mitchell (2018).

In the context of health, this phenomenon is called the “disability paradox” (Ubel et al., 2005). Ubel et al. (2005) coined this name for the phenomenon that people affected by disease judge themselves to be better off than one would expect from an outsider's perspective. One possible reason for this difference echoes Sen’s argument presented above: patients do not recognize how bad their health state is. One case that supports this argument can be found in a study by Smith et al. (2006), who studied the utility value of having a colostomy. Amongst other things, members of the general public, colostomy patients, and former colostomy patients were asked to conduct a TTO to value living with colostomy in this study. The result of this study was that former patients and members of the general public gave a similar low value of utility to living with a colostomy, while current patients gave a much higher utility value. In part due to adapted preferences, QALYs estimated with members of the general public are favored over those estimated with patients (Mitchell et al., 2015).

To conclude, desire fulfillment and happiness can indeed be seen as important for wellbeing and consequently should be used as an informational input for value judgements regarding the wellbeing of individuals. However, as *exclusive* sources of information, Sen argues that they are too limited to inform value assessments (Sen, 1985a, 1987). Indeed, Sen argues that the assessment of value requires information that goes beyond the happiness or desire fulfillment that individuals report (Sen, 1985a, 1987).

1.5 The capability approach

In light of his critique, Sen developed an alternative approach to assess the wellbeing of individuals. Sen argues that the assessment of wellbeing should be based on the capability of individuals to do or be, instead of their functionings, which are reflected

by individuals' beings and doings (Sen 1993b). Citing Sen, a capability "…reflects the alternative combinations of functionings the person can achieve, and from which he or she can choose one collection" (Sen 1993b).

One classic example by Sen to explain the difference between functioning and capability is the comparison between the following individuals: individual A is hungry due to famine and individual B is hungry due to following a diet. Based on the actual level of food intake and the level of hunger of these individuals it is impossible to decide who is better off. However, individual B chooses to be hungry, given that he or she chooses to follow a diet. Individual A does not have a choice, given that there is not enough food in the area where she or he lives. Because of the freedom that individual B has, Sen argues that he or she is better off (Sen, 1985a). In the capability approach, this freedom is called a "capability". The actual doings and beings of individuals, in the example above the actual food intake, are called "functionings". Assessing the capabilities of individuals thus leads to a more comprehensive evaluation of wellbeing than an evaluation that is limited to the assessment of functionings.

Extending the informational base from utility (in either of its three different understandings) to capability would lead to an improved assessment of wellbeing and would consequently improve the valuation process. As was mentioned, the value of a good in the choice understanding of utility is consequentialist. Amongst other things, this means that the utility value of a set of goods depends on its single goods, and not on the number of goods. In other words, regardless of the number of alternatives, the utility of a set of goods to an individual depends on its most preferred good. This is because each individual good has a specific utility value to an individual, and this value is not affected by the availability of other goods in a set. However, it could be argued that the ability to make a choice itself is also valuable. By extending the evaluation of individual wellbeing to the capabilities of individuals, the value of choice itself is also captured (Sen, 1985a, 1987).

One further advantage of the use of capability information over information about choice behavior is that it also captures the value of opportunities that might not benefit a particular individual directly (Sen, 1992c). To illustrate, it might be possible that an individual chooses one opportunity from a set of alternatives. In the choice understanding of utility, only this specific opportunity has value. The other opportunities

remain without value, given that no choice behavior could be observed which could be used to attribute value to the alternatives. However, within the capability approach, the value of these alternatives is acknowledged, given that having different opportunities itself is inherently valuable, irrespective if an individual values the ability to choose between one of them (Sen, 1992c).

Lastly, by focusing on the assessment of wellbeing in terms of capabilities, the adaptation problem is theoretically bypassed (Sen, 1992d). By evaluating the real opportunities that people have, one's assessment of wellbeing is not influenced by the happiness that an individual reports who has few opportunities. To illustrate, even when an individual with few opportunities is happy, from a capabilities perspective it is still possible to recognize that that individual has fewer opportunities and therefore a lower level of wellbeing than an individual who also reports being happy but with many opportunities (Sen, 1992d). Consequently, it is possible to distinguish between different levels of wellbeing of individuals, irrespective of the experienced level of wellbeing by the individuals themselves (Sen, 1992d).

1.5.1 Capability as a freedom

There are thus theoretical advantages to assessing the wellbeing of individuals in terms of capabilities. However, what kind of freedom is a capability? Sen conceptualized capability as a kind of "positive freedom" (Robeyns, 2017a). Positive freedom is a concept from the philosophy of liberty developed by Berlin (1969). Berlin distinguished two types of freedom: positive and negative freedom. According to the concept of "negative freedom", one can be considered free if nothing or nobody interferes with one's freedom (Berlin, 1969). According to the "positive freedom" concept, one can be considered free if he or she makes choices on his or her own, or, in the words of Robeyns, as being "one's master" and being able to make choices that reflect one's "true self" (Robeyns, 2017c).

The following example clarifies this difference (Carter, 2003). Imagine an individual who is traveling in a car to visit a school play from her niece, which is something that this individual greatly values. This individual arrives at a crossroads. Nothing blocks this individual from going left or right. Thus, nothing impairs the negative freedom of this individual. The individual is however addicted to smoking. When the individual turns left, then she will make it in time for her niece's play. However, when the individual

turns right, she will be able to buy cigarettes. If, due to her addiction, the individual decided to turn right and buy cigarettes, then it could be argued that her freedom is limited, given that this individual is not able to pursue what is actually worthwhile to her (Carter, 2003).

There are however two problems with Sen's conceptualization of capability as a positive freedom. The first problem is that positive freedom has been linked to paternalistic or even tyrannical regimes (Berlin, 1969; Robeyns, 2017c). In the context of positive freedom, a distinction can be made between choices that are "true" to one's self (e.g., visiting the niece's school play) and choices that are "untrue" to one's self (e.g., choosing to buy cigarettes). Consequently, it might be possible that an external entity claims to know individuals' "true selves" better than the individuals themselves. According to Berlin, this is what tyrannical regimes historically claimed, by appealing to "higher causes" that individuals were supposed to follow, which were unsurprisingly in line with the regimes' ideologies (Berlin, 1969; Robeyns, 2017c). This is inconsistent with Sen's understanding of capability (Robeyns, 2017c).

The second problem with understanding capability as a positive freedom is that according to Berlin, reductions in negative freedom do not necessarily lead to reductions in positive freedom (Berlin, 1969; Robeyns, 2017c). Taking the example earlier, if that individual was blocked from buying cigarettes, then that would not have affected that individual's freedom, given that the "true self" of that individual wanted to go left in the first place to visit the school play. In fact, it could be argued that by impairing the negative freedom of that individual, her positive freedom has increased, given that she will not be able to buy cigarettes again (Robeyns, 2017c). An example in the field of health might be a ban on participating in risky sports to reduce the number of injuries.

Sen however argued that in the context of capabilities, an impairment in negative freedom is always a violation of positive freedom (Robeyns, 2017c). As a consequence of these two problems, authors have made an effort to develop conceptualizations of capability that more closely reflect Sen's original ideas of freedom. An in-depth discussion about these different conceptualizations of capability goes beyond this dissertation and can be found in Robeyns (2017c). Instead, I follow the expert recommendation of Robeyns and define capability as an "option freedom". Option

freedom has been developed by Pettit (2003) as a theory of freedom. Option freedom is characterized by two elements: (1) options and (2) access to those options.

Pettit defined “options” as “the alternatives that an individual is in a position to realize” (Pettit, 2003). Options themselves can be characterized in terms of their quantity (i.e., the number of options) and their quality (i.e., the diversity between different options). Furthermore, options can differ in the way that they are objectively and subjectively significant (Pettit, 2003). Options that are objectively significant have the ability to affect the world. For example, the option of pressing a button on a remote with full batteries has a higher objective significance than the option of pressing a button on a remote with empty batteries. Consequently, one has more freedom with a remote with full batteries. Options that are subjectively significant carry meaning to an individual. For example, a sports fan would experience greater freedom when she or he can choose between tickets to different sports games as opposed to someone who does not like sports at all.

The “access to options” reflects an individual’s ability to realize options (Pettit, 2003). Access to options can be blocked or burdened (Pettit, 2003). When access to an option is blocked, it is not possible for an individual to realize that option. In the case of access being burdened, an individual would still be able to access an option, but with greater effort. The nature of these blocks or burdens can be objective or subjective. An example of an objective burden are laws that increase the difficulty for women to participate in political processes. Subjective blocks and burdens represent the perceived ability of individuals to access options. As an example, a woman could believe that she is an inferior manager compared to men, due to a belief in society that men are natural leaders. Objectively, this is not the case, but this woman might decide not to apply for management positions due to this subjective idea. In this case, the woman’s access to the option of becoming a manager is subjectively blocked.

Freedom understood as consisting of the elements “options” and “access to options” closely reflects the concept of capability that Sen implicitly uses and is further clarified in his writings in “Inequality Re-examined” (Robeyns, 2017c; Sen, 1992a). In particular, the notion that people are able to access options that are valuable to them is relevant in the context of the capability approach, since it acknowledges that burdens in access

to options result in a limitation to overall freedom. This is not reflected in the positive freedom understanding of capability by Sen (Robeyns, 2017c).

1.5.2 Capability lists

Another important question is what kind of capabilities should be selected to evaluate the wellbeing of individuals. Such a selection of capabilities is also called a capability list. A rough distinction can be made between two different approaches to developing capability lists: that of Sen and that of Nussbaum. Sen himself did not propose a definitive list of capabilities. Instead, he argues that a single list of capabilities is too inflexible, given that different people might value different capabilities, and that the type of capabilities that are valued might change over time. Furthermore, different lists might be developed for different uses (Sen, 2004). Therefore, Sen argues that a list of capabilities should be the product of a public discourse with people for which the list is developed (Sen, 2004).

Sen's decision to not develop a definitive list has prompted authors to develop lists themselves for various uses (Robeyns, 2003). Arguably the most influential list has been developed by Nussbaum (Nussbaum, 2003). Nussbaum proposed a list of ten different capabilities which she argues to be central for human wellbeing: length of life; health; bodily integrity; senses including imagination and thought; emotions; practical reasons; affiliations; other species; play; and, control over one's environment (Nussbaum, 2003). With this list, Nussbaum argues that it is possible to develop a list of capabilities that is relevant for all people. The list is intentionally developed to be on a highly abstract level and should be adapted to the local context when it is applied to, for example, wellbeing assessment (Nussbaum, 2003).

1.5.3 Adapted preferences in the capability approach

As mentioned in Section 1.2.1, one of the advantages of assessing wellbeing in terms of capabilities is that the adaptation problem is theoretically bypassed. Still, adapted preferences might pose a problem for empirical applications of the capability approach (Robeyns, 2017b). Robeyns (2017b) distinguishes two challenges. The first lies in the use of public reasoning to select capabilities for list construction. A systematically disadvantaged group might not select relevant capabilities on their list, since they might not be even aware of their relevance in the first place. An example of such a capability is women being able to follow education in repressed societies, which these women

might not consider important since they are expected to take care of the family. The second challenge relates to the perceived capabilities of individuals. Individuals might perceive to have less access to capabilities than objectively available to them, which could influence their choices. Individuals might then choose a suboptimal selection of functionings. However, identifying whether this selection of functionings is suboptimal is difficult for outsiders, since from an outside perspective the choice to not select more optimal functionings can be interpreted as a matter of agency of individuals that should be respected.

1.5.4 The capability approach in health economics

Given the theoretical advantages of using the capability approach for wellbeing assessment, it is unsurprising that authors have argued for its use in health economics to measure the effects of health technologies. Some insights from the capability approach might be particularly relevant for health economists. The critique of Sen on using information about choice behavior might be particularly relevant for the TTO and the SG, given that both methods utilize information about individuals' choice behavior to estimate values for health states (Sen, 1985a; Torrance, 1986).

An additional critique from proponents of the capability approach in health economics is that the current instruments that are used to assess value are limited in their scope. As was mentioned, the QALY is a measure that combines length of life and quality of life. Developed within the extra-welfarist context, this quality adjustment is a preference-adjusted score of a health state that reflects the HRQoL of an individual (Drummond et al., 2015b). Proponents of the capability approach argue that new medical technologies affect individuals' lives beyond health. Consequently, they argue for extending the informational base of assessing the effect of health technologies to include those elements that the people themselves consider to be important (Coast et al., 2008b; Lorgelly et al., 2010). Lastly, and perhaps most importantly, the benefit of applying the capability approach is that it forces health economists to reflect on their own theoretical assumptions and empirical methods, which implicitly or explicitly underlie their value judgements regarding what to measure. This could support the development of health economics into a discipline that bases its value judgements on principles that are more in line with that of society (Coast et al., 2008b).

1.5.5 Conclusion

In this section, I introduced various theories of wellbeing that can be used to assess the state of individuals' lives. Each of these theories argues for different elements to be key informational inputs for the assessment of wellbeing. However, the capability approach claims to offer the broadest, as well as the most appropriate informational base for the assessment of wellbeing. In health economics, this broad informational base has been used as a justification to develop new instruments to assess the effect of medical technologies on individuals' lives. Constructing such instruments is however challenging, due to the flexibility of the capability approach in terms of what kind of capabilities to include, as well as the multitude of different ways that concepts from the capability approach can be conceptualized. This results in researchers interpreting and operationalizing the capability in a number of different ways. Given the importance of instruments in health economic analysis and the subject of this thesis, I shortly introduce how they are developed in the next section.

1.6 Development of instruments

First, it is important to reflect on what the use of an instrument is. Instruments are used to measure constructs. In this context, a construct should be understood as a measurable psychological entity (for example a characteristic or an attribute). As such, a construct differs from "concepts" or "themes" (Gardner, 1996). To illustrate, an instrument that aims to evaluate how people experience the psychological entity "depression" might include the following items: "I feel depressed", "traveling to my psychiatric treatment is expensive" and "my family is affected by my depression". Conceptually and thematically each of these items is related to living with depression, however, one can imagine that responses to these items are not closely correlated with each other since each of them seem to reflect a different construct. This means that the sum scores of a scale with these items are difficult to interpret since it is unclear if the patient has financial problems, feels depressed, or whether her or his family is affected by the depression of their family member. Now imagine the following three items "I feel depressed", "I feel down", and "I feel sad". In this case, a sum score may give a reasonable insight into the emotional state of an individual, given that the items are closely associated with each other and reflect one construct: the emotional experience of sadness, which could be a symptom of depression (this example is inspired by Gardner (1996)).

1.6.1 Instrument development for survey research

How are instruments then developed? Many guides exist on the development of instruments (see Kyriazos and Stalikas (2018) for an overview of different guides). In this dissertation, the general guidelines developed by Boateng et al. (2018) are used. This guide was followed for two reasons. First, it provides guidance beyond immediate instrument development by addressing the scale evaluation phase. This particularly helps to provide context to Chapter 5 and Appendix Section 10.5 of the dissertation, in which various validation studies are presented that go beyond immediate scale development. Second, the guideline integrates technical guides as well as practical experience into a concise primer that broadly covers relevant steps in instrument development.

The guide by Boateng et al. (2018) consists of nine steps. The first three steps can generally be categorized as the qualitative part of instrument development. Step (1) is to clearly define the constructs that should be measured. This is arguably the most important step in instrument development since a good definition establishes which constructs an instrument should measure. Based on this definition, items can be generated to measure these constructs. Step (2) is to assess the content validity of the instrument. The content validity of an instrument is determined by assessing whether the items that are linked to a certain construct are representing the construct of interest (DeVellis, 2017e). In step (3) of instrument development, the items have to be qualitatively pre-tested. Such pre-tests aim to establish two things (Lenzner et al., 2016): One, to study whether the items themselves reflect the construct of interest when administered to participants. Two, to establish if the participants' responses to the items are in line with their evaluation of the construct of interest. These pre-tests can be done with a variety of different methods, but they essentially all share that participants from the target population of an instrument are asked to complete the instrument, where special attention is paid to the way participants interpret the items to assess whether the items measure the construct of interest.

Step (4) is to administer a set of items from which an instrument can be formed to a sample of participants. The sample should be sufficiently heterogeneous, to test how the instrument functions in different types of groups within the target population (Clark and Watson, 1995). For example, when developing an instrument that aims to support

the diagnosis of depression, the instrument must function well in participants with different types of symptoms. Also, the sample has to be sufficiently large. In the context of instrument development, it is difficult to say how large of a sample is needed (DeVellis, 2017a). Such an estimation should take into consideration both the absolute number of participants that is sufficient for analysis, as well as the relative number that is needed. An estimation of the relative number of participants that are needed depends on, for example, the statistical procedure that is being used for the analysis of the data, the number of parameters that need to be estimated, or the statistical properties of individual items. With an increasing sample size, the need to account for these contextual factors diminishes. To illustrate, Tinsley and Tinsley (1987) note that a ratio of 5 to 10 participants per item is required for analysis, with an upper limit of 300. This ratio can be relaxed once the sample size is larger than 300. Comrey (1988) suggested that a sample size of 200 is sufficient for most forms of factor analysis conducted on a dataset with fewer than 40 items.

Step (5) is to reduce the number of items that have been developed in steps (1) and (2). One strategy is to study the “item discrimination” and “item difficulty” properties. Essentially, an item’s discrimination properties give researchers information on whether an item correctly differentiates between different levels of the construct of interest in a respondent. In classical test theory, item-rest correlations can be used to establish an item’s discrimination property. This is a correlation between an item of interest and the total score of a scale minus the item of interest (Bechger et al., 2003). An item’s difficulty represents the ability of an item to produce information about a respondent’s state in the context of a certain construct, in light of the actual state of the respondent (Hambleton and Jones, 1993). To illustrate this one can imagine an item that is developed to assess the physical ability of an individual. This item covers the speed at which a respondent can run. In a sample of healthy individuals, is to be expected that this item differentiates between different levels of physical ability. In a sample with bedridden respondents, this item would however not yield much information, since not a single respondent can run. In the latter case, the item is considered to be too difficult for this sample.

In step (6) a factor analysis is conducted. One use of factor analysis is to help researchers understand how a limited number of latent variables predict the responses to a larger set of items (DeVellis, 2017a). This is done by developing a factor model,

which consists of factors (the latent variables) and items linked to those factors. In this model, responses to items are predicted by the factor. The model that captures the relationship between factors and their items is called a measurement model (DeVellis, 2017d). The relationship between the factors themselves (and possible observable variables influencing the factors) is called the structural model (Kline, 2011c).

Key statistics of a factor model are the number of factors and the factor loadings. By studying the number of factors and their correlations, researchers can get an insight into the latent structure of these factors, which in the context of instrument development means that researchers get an insight into the number and type of psychological constructs that are associated with observable responses to items (DeVellis, 2017a). Factor loadings are regression coefficients and represent the strength of a relationship between the factor and an item (DeVellis, 2017a). The relationship between a factor and its associated items informs researchers about the type of construct that a factor represents. An example might be three items that ask about experiencing pleasure, being content, and being cheerful. When developing a factor model with these three items, one can expect the factor loadings to be high on one factor, which informs researchers that this factor covers the experience of happiness.

In general terms, two different types of analytical tools can be differentiated from each other in factor analysis: exploratory factor analysis and confirmatory factor analysis (DeVellis, 2017a). In an exploratory factor analysis, researchers make no a priori assumptions regarding the relationships between individual items and potential factors. With statistical tools, a model is developed which effectively informs how a limited number of latent variables (or factors) can optimally predict a covariance structure. A confirmatory factor analysis is conducted with an a priori idea of the factor structure, as well as the relationship between items and factors (DeVellis, 2017a). As such, confirmatory factor analysis can be used for hypothesis testing to confirm whether certain constructs can be measured. For instrument development, it is important to confirm that items on a single scale are unidimensional instead of multidimensional (DeVellis, 2017a). An item is considered unidimensional when it loads on one factor and its error terms are independent. Unidimensionality facilitates the interpretation of scale scores since all the items on a scale reflect one construct. This is not possible with a scale that consists of multidimensional items, since it is unclear what construct

has exactly influenced responses to an item, thus complicating the interpretation of scale scores.

Step (7) is to conduct further dimensionality tests. Several different tests can be conducted to test for dimensionality. Boateng et al. (2018) discuss several methods, one of them being measurement invariance testing. In measurement invariance testing, the aim is to study whether the same measurement model or structural model can be applied to different groups. Successfully doing so indicates that the items of an instrument are interpreted similarly in different groups and are thus unidimensional, in the sense that differences irrelevant in group characteristics do not influence responses to items. Further information about measurement invariance testing can be found in Chapter 5.

Step (8) is to test the reliability of the instrument. The reliability of an instrument is the extent to which an instrument “behaves” similarly under similar or identical circumstances (DeVellis, 2017c). One measure of reliability is internal consistency. Internal consistency reflects the degree of the interrelatedness of items on a scale (DeVellis, 2017c). One measure of internal consistency is Cronbach’s alpha, which is calculated by averaging the correlations of all pairs of items on a scale, adjusted for the number of items on that scale (DeVellis, 2017c). According to Kline (2011a), a value higher than 0.90 can be considered excellent, values around 0.80 as very good, and values around 0.70 as adequate. Also, other tests of reliability exist. An overview can be found in DeVellis (2017c). Step (9) is to further assess the validity of the instrument (Boateng et al., 2018). This is an ongoing process, where various types of validity are continuously assessed when applying the instrument to new groups of participants (DeVellis, 2017e). Information from validity tests can be used to assess the strengths and weaknesses of an instrument in different contexts (DeVellis, 2017e).

Notable forms of validity are construct validity and criterion validity. An instrument is considered to have construct validity when it measures what it is intended to measure (DeVellis, 2017e). Typically, this is studied by establishing whether theoretical relationships between variables can empirically be identified. In this case, it is expected that variables that theoretically measure similar elements are correlated with each other (convergent validity) (DeVellis, 2017e). Additionally, variables that theoretically measure different elements should not be correlated with each other (discriminant

validity). Criterion validity represents the idea that a measure empirically correlates with a “golden standard”. This can be studied by testing the correlation with a criterion variable that was measured in the past (postdictive validity), is measured jointly (concurrent validity), or is measured in the future (predictive validity).

It should be noted that the steps presented above presume the use of classical test theory in instrument development. Alternatively, item response theory can be used to develop instruments. Both methods differ in their approach to psychometric analysis. Analysis in classical test theory is primarily based on scale-level information. It is assumed that scale scores reflect a combination of a true score directly linked to the construct of interest and a random error, which explains the difference between the observed score as measured by a scale and the true score. In other words, observed scale scores are interpreted as a reflection of the construct of interest with an error component (Fan, 1998). This focus on scale-level information leads however to a problem. Item statistics (e.g., item discrimination and item difficulty) depend on the manifestation of a construct in a sample. Person statistics, such as observed scores scale scores in a sample, depend on item statistics (Fan, 1998). In other words, a certain response to an item does not only reflect the construct of interest, but also the item discrimination and the item difficulty within a certain population. This circular dependence on the sample makes a comparison of observed scores and item statistics across different samples challenging (De Champlain, 2010). Therefore, test scores can only be compared in populations in which the construct of interest manifests similarly (De Champlain, 2010). It should however be noted that various methods have been developed to address this problem (Fan, 1998).

In item response theory, items are modeled according to a probabilistic model that reflects the probability that an individual gives a certain response to an item, given the degree that an individual is under influence or affected by a certain construct (Fan, 1998). As the name suggests, in item response theory primarily information about individual items is used. Furthermore, psychometric analyses in item response theory are, in theory, not sample dependent, since item characteristics are estimated on an item level instead of being derived from observed scores (Fan, 1998). However, this comes at the cost of additional assumptions about the relationship between constructs and item responses, which results in models with a larger number of parameters and additional assumptions regarding the functional form of item responses (Fan, 1998).

In this dissertation, classical test theory will be used for psychometric analysis and instrument development. This is done for two reasons. One, an advantage of classical test theory is that psychometric models in classical test theory require fewer theoretical assumptions regarding the relationship between items and constructs compared to item response theory (Fan, 1998). Second, various studies have shown that scales developed using either item response theory or classical test theory are mostly similar (Fan, 1998; Macdonald and Paunonen, 2002; Progar et al., 2008). Due to the marginal additional benefit of item response theory for scale development that comes at the cost of stronger assumptions and more complex modeling of item responses, I decided to use classical test theory for the work in this dissertation.

1.6.2 Instrument development in the capability approach

Essentially, the content of a capability instrument is a capability list (i.e., such as the list of Nussbaum, see Section 1.5.2), covering capabilities and functionings that are relevant to the research at hand (Robeyns, 2005). Such a list is inherently normative since it prescribes which components of wellbeing are relevant for people (Robeyns, 2005). Studies with instruments that are used in health economics illustrate why this is important in practice. These instruments have been shown to differ in their sensitivity to particular health problems due to differences in their content (Khan and Richardson, 2018; Richardson et al., 2015; Richardson et al., 2016). This has the effect that depending on the choice of instrument, the estimated effect of a health technology on a certain health problem might change by more than 100% (Richardson et al., 2016). Consequently, the question of which content to include in instruments is not only a value judgement about what is important in individuals' lives but also affects the value assessment of health technologies.

To ensure that capability lists are developed appropriately, Robeyns (2005) developed four criteria. The first criterion is that the capability list should be explicitly formulated, discussed, and defended. It is important that the list is developed before the analysis. This will help identify the limitations of the analysis in terms of which capabilities or functionings should have been analyzed, but were not in the end because of a lack of variables in the dataset. The second criterion is that the methods that are used to develop the list should be justified. The justification should explain why the chosen method was appropriate to the research context for which the capability list is

developed. The third criterion is that when a capability list is developed for empirical research, the list should be developed in two stages: in the first stage, an “ideal” list should be developed. In the second stage, the list can be adapted to the limitations that are often inherent to empirical research. Examples of such limitations are that the data might be limited, or there might be socio-economic or political constraints. According to Robeyns (2005), it is important to distinguish between these two levels, since limitations in data collection might disappear over time. The fourth and last criterion is that the capability list should be exhaustive and non-reductive. In other words, the list should contain all the elements that are important for the specific use of that capability list.

When considering the guidance provided by Boateng et al. (2018) and the procedural criteria by Robeyns (2005), it is important to note that there is a strong parallel: both point to the importance of conceptual work as the core of instrument development. Frameworks or concepts guide researchers to focus on a particular aspect of a problem or a research question, while other aspects are moved to the periphery (Bordage, 2009). For example, in the context of wellbeing research, authors who understand wellbeing as a form of pleasure may ignore other aspects of wellbeing that may be relevant for wellbeing assessment. Selecting an appropriate framework thus supports researchers to reach a richer and deeper understanding of the question at hand (Bordage, 2009). This is particularly important for instruments that are used to evaluate individuals’ lives to inform policy-making, given their potential influence on the allocation of scarce resources.

1.7 Conclusion and objectives of this dissertation

Theories of value, theories of the good, and theories of wellbeing are interlinked with each other. In applications of these theories in assessments of how well-off individuals are, value judgements are made about what kind of information should be used. In conventional health economics, these value judgements are based on welfarist and extra-welfarist theory. Instruments in health economics that are based on these theories measure important aspects of health-related wellbeing.

However, insights from the capability approach suggest that instruments that are based on welfarism or extra-welfarism might not capture all the elements that are important for individuals. Proponents of the capability approach argue that a broader

perspective is needed to measure and consequently value the effect of health technologies. Indeed, they argue that health technologies should be assessed in terms of their effect on the capabilities of individuals. These capabilities reflect the freedom of individuals to do or achieve things that they value.

Developing instruments based on the concept of capability is not without challenges. One important element of instrument development is to have a clear definition of the concept that needs to be measured. The definition of capability as proposed by Sen is however not very clear in terms of the kind of freedom that it represents. As such, the concept of capability itself can be understood in several different ways. Consequently, researchers might differ in the way that they operationalize the concept of capability, which leads to different choices in terms of the type of content that they include in the instruments that they develop.

Given the normative implications of including or excluding certain types of content from an instrument that is used to assess wellbeing, it is important to understand the relationship between concepts of capability and how they are operationalized in capability instruments. By studying this relationship, there might be further lessons in how to use the capability approach for the development of instruments.

Therefore, the objective of this dissertation is to study and advance the application of the capability approach in the development of instruments that can be used to assess the effect of health technologies. This was done in four steps:

1. Study how the capability approach is currently conceptualized and applied in instrument development. This is the subject of Chapter 2.
2. Develop a theoretical framework that is based on the concept of option freedom that can be used to develop an instrument. This is the subject of Chapter 3.
3. Study whether the themes of this theoretical framework can be operationalized as constructs in an instrument. This will be the subject of Chapter 4.
4. Study how psychometric methods can be used to explore whether responses to instruments are influenced by adapted preferences. This is the subject of Chapter 5.

2 CAPABILITY INSTRUMENTS IN THE FIELD OF HEALTH RESEARCH

2.1 Background

In the context of health economics, a number of capability instruments have been developed that can be used to assess the effect of health technologies. Mitchell et al. (2017) identified four such instruments. The developers of these instruments used different methods to operationalize the capability approach for wellbeing assessment, resulting in a wide variety of content (Kinghorn, 2015; Mitchell et al., 2017). As highlighted in the introduction, the content of the instruments determines their sensitivity to certain changes in wellbeing. Thus, because of these differences in the content of capability instruments, it is possible that they differ in their ability to assess capability wellbeing in particular populations.

Kinghorn (2015) notes that one reason for this variation is related to how the developers of the instruments interpret the capability approach. The term “developers” refers to the group of researchers who have created a particular instrument. A thorough review about how researchers conceptualize capability and how these concepts have been operationalized in the form of capability instruments is lacking.

Therefore, the aim of the narrative literature review presented in this chapter is to study how researchers operationalized their understanding of capability in instruments that can be used to assess wellbeing within the field of health. This study was conducted through three sub-aims. The first sub-aim is to analyze what kind of concepts of capability were used by researchers who developed instruments and why these concepts were chosen. The second sub-aim is to study how the content of the instruments relates to the concepts of capability that were chosen by the researchers. The third sub-aim is to compare the content of the capability instruments to a more comprehensive concept of capability: option freedom. This last step might highlight the potential strengths and weaknesses of instruments in relation to the measurement of freedom.

2.2 Methods

2.2.1 Pearl growing method

The identification of capability instruments in databases is challenging since the word “capability” itself leads to an unmanageable amount of hits. Therefore, a search method called “comprehensive pearl growing” was used to identify capability instruments (Schlosser et al., 2006). With comprehensive pearl growing, initial key references (the “pearls”) are identified, which are used to identify further papers that are of interest by searching for the papers that reference the pearl. This results in the first wave of papers. If studies that were of interest were identified, then these new studies were used to look for further studies, which formed their own consecutive waves. This process was repeated until no new articles of interest are identified.

PubMed and Web of Science were used to identify the initial pearls, with the search string ("Capability Approach") AND ("Measure" OR "Outcome" OR "Empirical" OR "Index" OR "Operationalization" OR "Instrument" OR "Questionnaire" OR "Attributes" OR "Domains" OR "Evaluation"). These databases were also used in subsequent waves to identify capability instruments. Abstracts that mentioned anything related to the capability approach or the assessment of wellbeing with broader dimensions were included. In July and August 2018, the first literature search was conducted. The literature search was updated in April 2021. The abstracts of the first literature search were independently screened by me and KHV, a colleague from the division of health economics. The new abstracts of the updated search were screened by me, with a random one-fifth selection of the new abstracts checked by KHV. There were no differences between KHV and me regarding the inclusion of these abstracts.

2.2.2 Study selection

Once an article was included on the basis of its abstract, the full paper was read by me. In this process, any kind of self-reported instruments developed for the evaluation of health technologies on wellbeing in terms of capability was identified and indexed. Adaptations of capability instruments or translations were used to identify the original capability instrument on which they were based. Articles were identified as pearls for subsequent waves if they mentioned any kind of self-report capability instrument that can be used to assess the effect of health technologies.

2.2.3 Method of analysis

The identified capability instruments were used to search for articles that explained how the content of the instruments was developed (hereafter called “development papers”). Development papers contain information about how the researchers who developed the capability instruments understood capability and explain how the content of the instruments was generated. These development papers were analyzed in three stages. In stage 1, the definition of capability that was used for instrument development was identified, as well as the reasons for using the capability approach as a framework for the development of an instrument. In stage 2, the different domains of capability wellbeing of the instruments were compared to the definition of capability used by the developers of the instruments. In stage 3, the content of instruments was analyzed to study in how far this content reflects the concept of capability.

To facilitate the analysis in stage 3, I decided to compare the content of the identified instruments to a more comprehensive and more precise concept framework of capability: the concept of option freedom. The concept of option freedom provides a clear and rich conceptualization of capability. This makes this concept useful to study the content of existing capability instruments in order to identify which aspects of capability are being measured (see Section 1.5.1 for an explanation about the concept of option freedom).

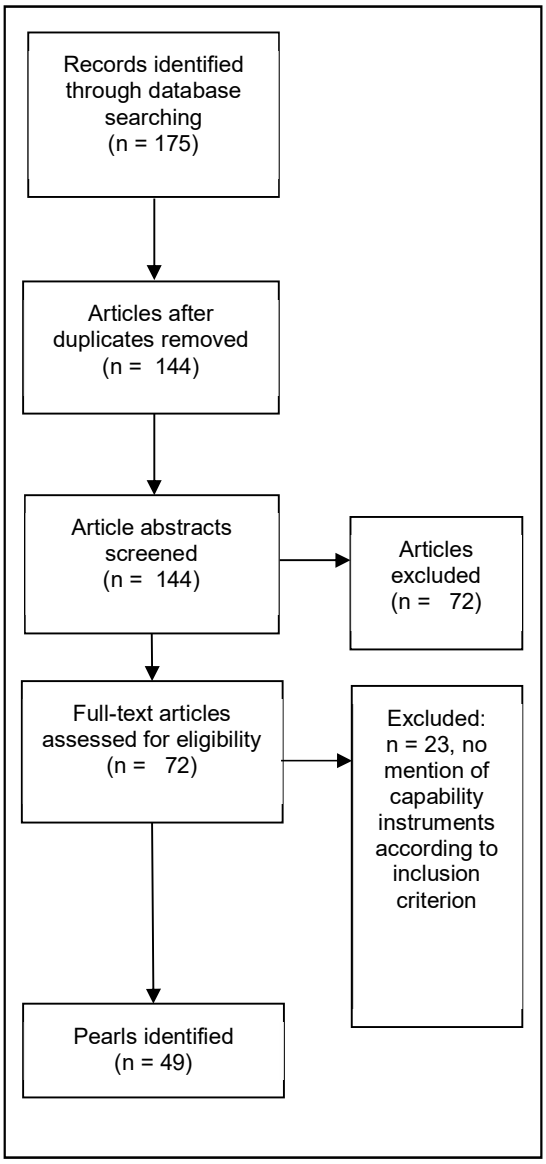
2.3 Results

2.3.1 Literature search and instrument identification.

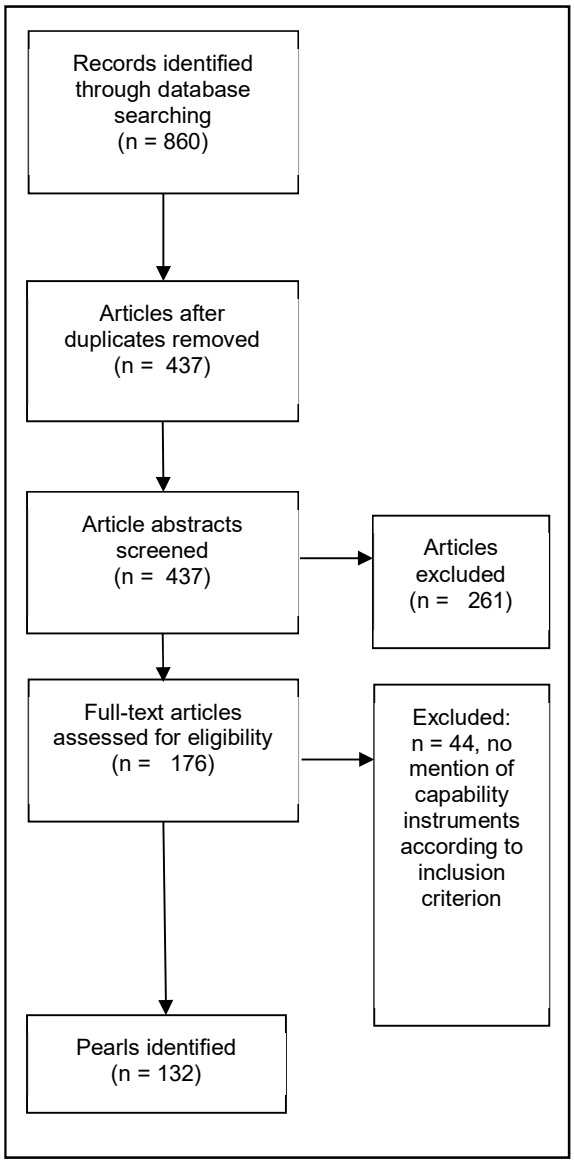
Figure 1 shows the flow diagram of the pearl growing search strategy per wave. Twelve capability instruments were identified in eight waves. These instruments and associated descriptive information are presented in
Table 1. Nine of the instruments were identified in wave one, and the remaining three in wave two. More detailed descriptive information can be found in Appendix Table 1.

Figure 1. Flow diagram of literature search

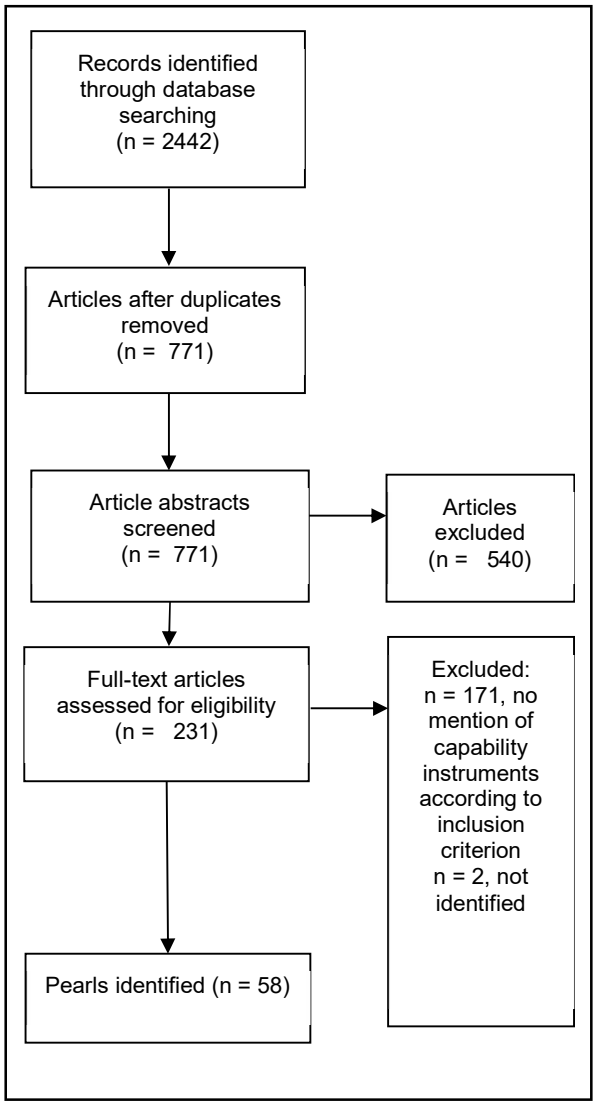
A. Wave 1



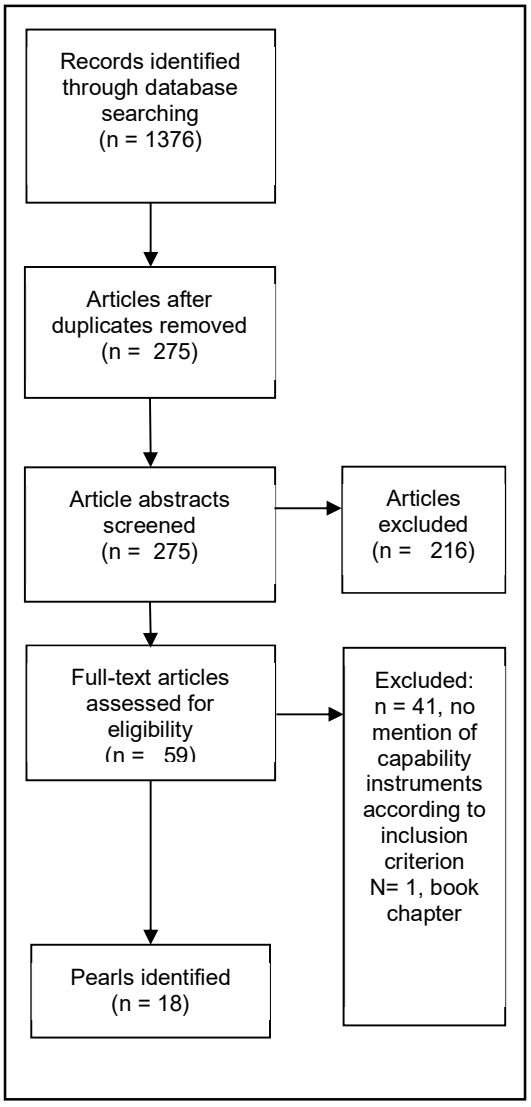
B. Wave 2



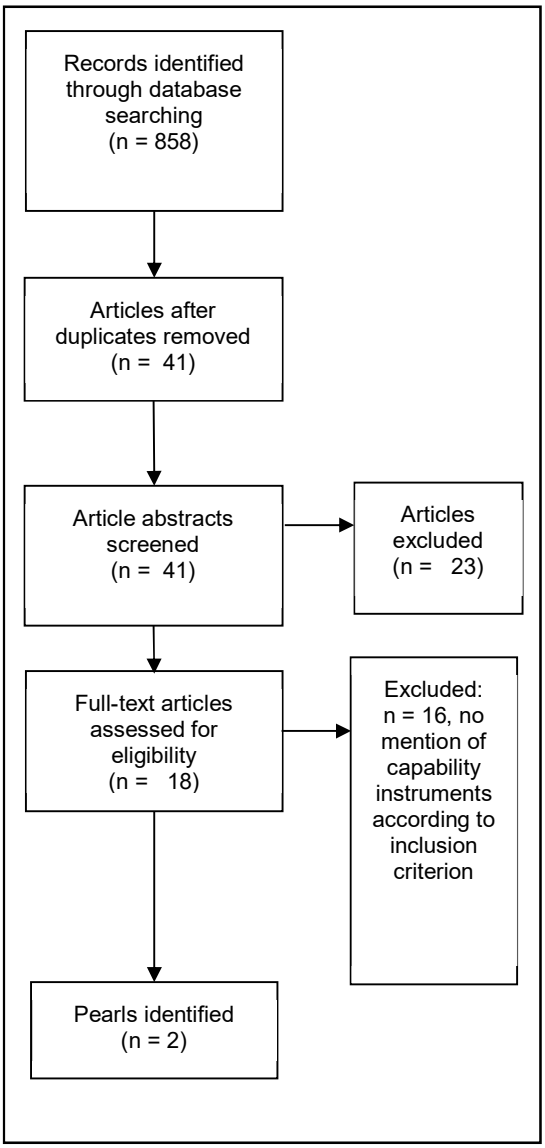
C. Wave 3



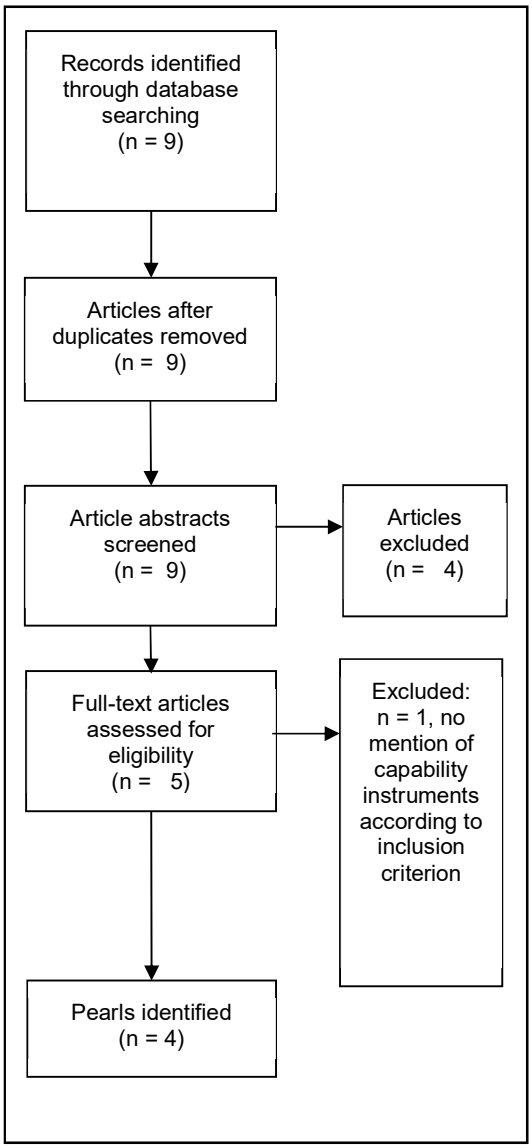
D. Wave 4



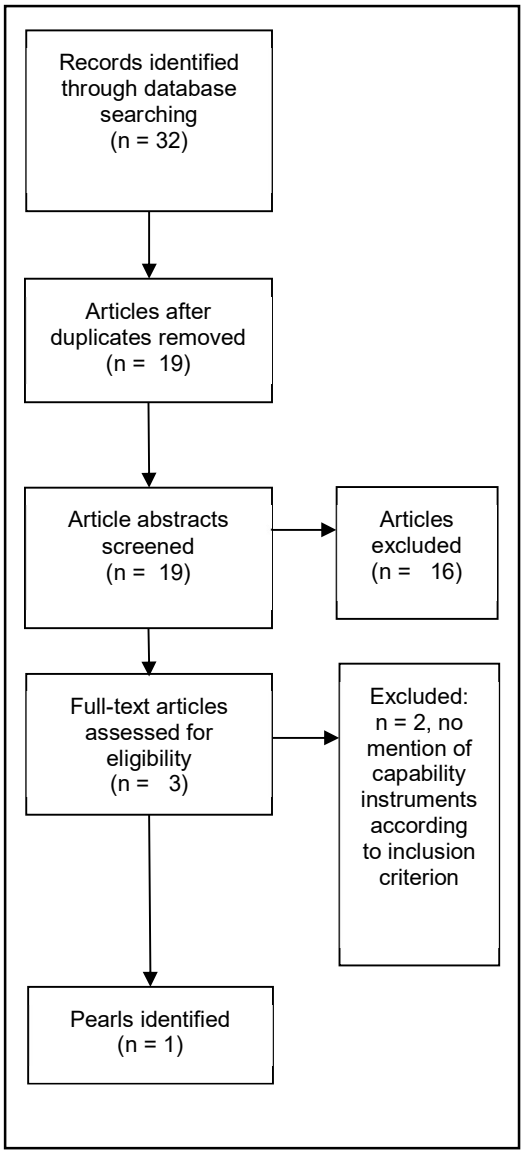
E. Wave 5



F. Wave 6



G. Wave 7



H. Wave 8

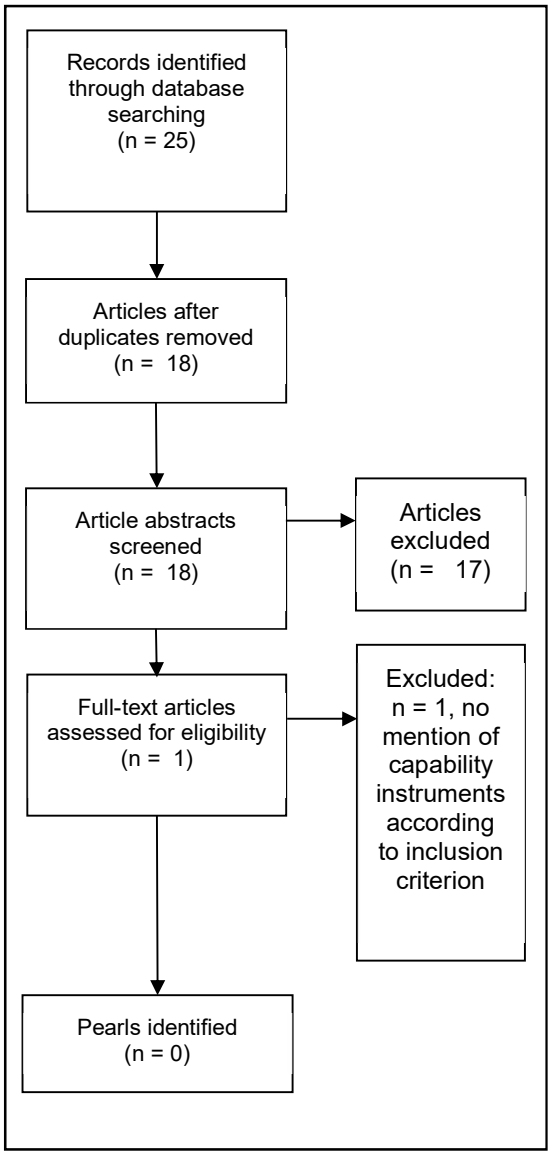


Figure published previously in Ubels et al. (2022b).

The identified instruments were in different stages of development. Some of the instruments were in an early, qualitative phase of instrument development (Kibel and Vanstone, 2017). Others have been psychometrically validated and used in a variety of settings, such as the ICEpop CAPability measure for Older people ICECAP-O (Grewal et al., 2006). Publications that explain the generation of content for capability instruments with rich qualitative data were identified for seven instruments at the time of writing this dissertation (Al-Janabi et al., 2012; Engström et al., 2016; Greco et al., 2015; Grewal et al., 2006; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Sutton and Coast, 2014). Articles with rich qualitative data contain detailed information about the qualitative process of developing the content for capability instruments. These articles

present the resulting qualitative framework that forms the basis of capability instruments with supportive illustrative quotes from participants and detailed explanation from the developers of instruments how they interpreted the qualitative data.

Table 1. Descriptive information of identified capability instruments

Instrument	Author(s)	Location	For which population the instrument is (being) developed to assess wellbeing
ASCOT	Netten et al. (2012)	England	Adults in contact with social care services
CALY	Månsdotter et al. (Månsdotter et al., 2017; Månsdotter et al., 2020)	Sweden	General population
Capability-based questionnaire for assessing well-being in patients with chronic pain	Kinghorn et al. (2015)	England	People affected by chronic pain
Child – and parent report questionnaire to explore capability of deaf children wearing a cochlear implant	Rijke et al. (2019)	Netherlands	Deaf children wearing a cochlear implant
Diabetes specific instrument for measuring patient reported outcomes and experiences in the Swedish National Diabetes Register	Engström et al. (Engström et al., 2018; Engström et al., 2016)	Sweden	Adults affected by diabetes
ICECAP-A	Al-Janabi et al. (2012)	England	General population
ICECAP-SCM	Sutton and Coast (2014)	England	People at the end of their lives
ICECAP-O	Grewal et al. (2006)	England	People of age 65+
Non-invasive prenatal testing related capability wellbeing	Kibel and Vanstone (2017)	Canada	Adult women
OCAP-18	Lorgelly et al. (Lorgelly et al., 2008; Lorgelly et al., 2015)	Scotland	General population
OxCAP-MH	Simon et al. (2013)	England	Adults affected by mental health problems
Women's capability index	Greco et al. (Greco, 2013; Greco et al., 2015)	Rural Malawi	Adult women

Table published previously in Ubels et al. (2022b).

2.3.2 Stage 1: Used concepts of capability for instrument development.

Table 2 gives an overview of the definitions of capability that have been used in the development papers and the main justification for why the developers chose to base their instrument on the capability approach. Researchers chose the capability approach as an a priori framework for instrument development in ten instruments (Al-Janabi et al., 2012; Engström et al., 2016; Greco et al., 2015; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Lorgelly et al., 2015; Månsdotter et al., 2017; Netten et al., 2012; Rijke et al., 2019; Simon et al., 2013). The main reason reported for choosing the capability approach as an a priori framework was its broad informational base in the context of wellbeing assessment. In this context, in six of the development papers it is explicitly mentioned that one of the strengths of the approach is that the assessment of wellbeing is conducted with broad domains that are of importance to individuals (Al-Janabi et al., 2012; Engström et al., 2016; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Månsdotter et al., 2017; Netten et al., 2012). In four development papers, it is argued that the strength of the approach lies in its argument for assessing wellbeing in terms of the freedoms of individuals (Kinghorn et al., 2015; Lorgelly et al., 2015; Netten et al., 2012; Rijke et al., 2019).

In the development paper of the ICECAP-O (Grewal et al., 2006) researchers explain that the capability approach is used a-posteriori as a framework to interpret the findings of the qualitative study. In the development paper of one other instrument, the ICECAP Supportive Care Measure (ICECAP-SCM), the capability approach is not mentioned at all (Sutton and Coast, 2014). Still, the instrument is part of a collection of capability instruments and was therefore included in the analysis presented in this chapter.

Except for the ICECAP-SCM, the concept of capability is explicitly understood as a kind of freedom in the development papers (Al-Janabi et al., 2012; Engström et al., 2018; Engström et al., 2016; Greco et al., 2015; Grewal et al., 2006; Kinghorn et al., 2015; Lorgelly et al., 2015; Månsdotter et al., 2017; Netten et al., 2012; Rijke et al., 2019; Simon et al., 2013). In these papers, freedom is described as the kind of opportunities individuals have, what individuals are “able to achieve” and what individuals “can do”. These definitions are based on Sen’s concept of capability (Sen, 1993a).

Table 2. Developers' definitions of capability and justification for why the capability approach was chosen as a framework per instrument

Instrument	Definition of capability used	Reason for choosing the capability approach as an a priori framework	Explicit argument for broader domains (bold)	Explicit argument for measuring freedom (italic)
ASCOT	"Sen argues that utility (or experienced QoL [quality of life] derived from functionings) is not the sole object of value; rather, it is capability – understood as the substantive opportunities an individual has to be, or to do, a range of things – that is the prime object of value."	"… *focus on choice and control encourages us to aim to measure what people can do, rather than what they actually do,* **across all aspects of SCRQoL** [social care related quality of life]."	X	X
CALY	"…according to Sen, the most important information to consider is capabilities, which refer to the opportunities to achieve a flourishing life according to an individual's own wishes"	"**Since public health interventions may impact other well-being components besides health and since social welfare policy and reform (education, labour market, social insurance, etc.) may also affect lifetime health**, it seems meaningful to establish a summary measure of capabilities."	X	-
Capability-based questionnaire for assessing well-being in patients with chronic pain	"The capability of a person is the alternative combinations of functionings the person can achieve, and from which he or she can choose one combination (Ibid.)."	"One strength of this approach, *which focuses on the freedom and ability of individuals to lead a life that they have reason to value, is its wide informational base*. **This can incorporate more of what is important to patients** and allow for the evaluation of a broader range of interventions."	X	X
Child – and parent report questionnaire to explore capability of deaf children wearing a	"…capability extends beyond an individual's actual functioning by asking what range of valued activities and modes of being are available to him."	". … In terms of their post-implant performance on hearing and speech tests, these children can generally achieve levels that are close to those of their normal-hearing peers [references]. Also in terms	-	X

Instrument	Definition of capability used	Reason for choosing the capability approach as an a priori framework	Explicit argument for broader domains (bold)	Explicit argument for measuring freedom (italic)
cochlear implant		of self-reported quality of life, their scores tend not to differ from those obtained in their peers [references]. However, performance scores on standardised hearing and speech tests may merely predict poor performance in day-to-day conditions,… while self-reported quality of life measures may be confounded by the response shift phenomenon, i.e., the respondents' adaptation to their (new) living conditions [references]. *The assessment of capability could then reveal whether, in spite of the cochlear implant and subsequent rehabilitation, children still experience constraints in pursuing their aspirations in terms of achievements and modes of being*."		
Diabetes specific instrument for measuring patient reported outcomes and experiences in the Swedish National Diabetes Register	"According to Sen, evaluation of the quality of life should focus on what individuals can do (capabilities) in relation to what they value as important in life rather than what they in fact do (functionings)."	"Sen's capability approach, which was used as a framework in this study, provides a general frame of thought and **urges that context and specific purpose need to be taken into account when selecting what aspects to evaluate."**	X	-
ICECAP-A	"The approach advocates assessing capability (what an individual can do) rather than functioning (what they actually do) to avoid imposing a particular idea of what a good life constitutes	[following the definition] "Whilst the capability approach was pioneered in human development research, focusing on basic capabilities such as being able to have shelter and being able to be nourished, **there is**	X	-

Instrument	Definition of capability used	Reason for choosing the capability approach as an a priori framework	Explicit argument for broader domains (bold)	Explicit argument for measuring freedom (italic)
	and to reflect the importance of freedom to choose.”	**recognition that measuring more complex capabilities can be useful for public policy**."		
ICECAP-SCM	Not given; part of ICECAP paper series	Not given; part of ICECAP paper series	NA	NA
ICECAP-O	"the extent to which a person is able to function in a particular way, whether or not he or she chooses to do so"	Capability approach used a-posteriori to interpret results	NA	NA
Non-invasive prenatal testing related capability wellbeing	"Its central normative proposition is that wellbeing assessments should be based on “what people can do” (their capabilities) as opposed to “what they actually do ” (their functionings)”	"**A capabilities approach suggests ways in which people might value NIPT [noninvasive prenatal testing] that go beyond clinical outcomes or quality of life**."	X	-
OCAP-18	"The capability approach suggests that wellbeing should be measured not according to what individuals actually do (functionings) but what they can do (capabilities)."	"Of interest in its application to public health is the evaluation space; it diverges from narrow utility space, which is concerned with the pleasure obtained from the consumption on goods and services, and *instead encapsulates an informational space, where evaluative judgements occur according to an individual’s freedom."*	-	X
OxCAP-MH	“Sen argues that outcomes (functional utilities) should not be the sole object of welfare assessments and that capabilities (things that people are free to do or be) should also be included in the overall assessment of a person’s wellbeing.”	"...more and more health economists and social scientists agree that the capabilities framework has the potential to offer a richer theoretical evaluative space compared with the traditional QALY approach and may have particular strengths when assessing complex interventions in social care and public health "	-	-

Instrument	Definition of capability used	Reason for choosing the capability approach as an a priori framework	Explicit argument for broader domains (bold)	Explicit argument for measuring freedom (italic)
Women's capability index	"… the abilities to achieve those "beings and doings" that people have reason to value in life."	"The Capability framework distinguishes itself from other conventional approaches, which have a narrower evaluative space, such as utility, income or basic needs. In order to improve people's quality of life, social and public policy should therefore aim to protect, restore and expand people's capabilities."	-	-

Text in bold represents an explicit argument for broader domains. Text in italic represents an explicit argument for the assessment of freedom.
Text in the following brackets "[ ]" has been added by me to further clarify aspects or signify where I changed the original text.
Table published previously in Ubels et al. (2022b).

2.3.3 Stage 2: Operationalizing of capability concept into content

Table 3 presents the domains that are included in the identified capability instruments, categorized around general themes. The researchers used different qualitative methods to develop domains and items for the capability instruments. The developers of five instruments started with an a priori idea of what kind of content should be included in a capability instrument. Of these five instruments, the developers of two instruments used Nussbaum's list of capabilities as an a priori source of domains for their qualitative studies. In these studies, Nussbaum's list was used to identify and further develop domains and items from an earlier instrument through focus group discussions (Lorgelly et al., 2008; Lorgelly et al., 2015; Simon et al., 2013). Developers of another instrument used Nussbaum's list to guide the secondary analysis of semi-structured interview data (Kibel and Vanstone, 2017). The researchers behind the Capability-Adjusted Life-Year (CALY) instrument used a report issued by the Swedish government as an a priori list of capabilities to develop their instrument (Månsdotter et al., 2020). The Adult Social Care Outcomes Toolkit (ASCOT) is the last instrument that is based on an a priori idea about what kind of content should be included (Netten et al., 2012). The ASCOT is a further development of an earlier non-capability approach-inspired instrument (Netten et al., 2012).

The developers of the seven instruments did not start with a predetermined list of capabilities. Domains for the ICECAP-A, the ICECAP-O, the ICECAP-SCM, and the capability-based diabetes questionnaire were identified with semi-structured interviews (Al-Janabi et al., 2012; Engström et al., 2016; Grewal et al., 2006; Sutton and Coast, 2014). For another instrument, the Women´s Capability Index (WCI), the researchers developed content with semi-structured focus group discussions (Greco et al., 2015). The developers of the capability-based questionnaire for assessing well-being in patients with chronic pain conducted both focus group discussions and interviews to develop the content for their instrument (Kinghorn et al., 2015). The developers of the child – and parent-report questionnaire to explore the capability of deaf children wearing cochlear implants utilized literature, conversations with parents of children with cochlear implants, and input from cochlear implementation experts to produce content (Rijke et al., 2019).

Another observation was that the twelve instruments were developed for use in different populations in different contexts. This in part explained the large variation in the content of instruments. For example, one instrument was specifically developed to measure those aspects that Swedish diabetes patients consider to be important for their lives (Engström et al., 2016). This resulted in very specific domains and items that are relevant to this patient group (see Table 3). Alternatively, some instruments consisted of content that is relevant to a broader public, such as the general population of England (Al-Janabi et al., 2012). Another reason for the variation in content between the capability instruments was that developers made different choices with respect to the level of abstraction of domains and items. For example, the domain “food and drink” from the ASCOT is very concrete (Netten et al., 2012). This can be contrasted with abstract domains of other instruments, such as “respect and identity” from the capability-based questionnaire for assessing well-being in patients with chronic pain (Kinghorn et al., 2015).

Table 3. Description of how capability was measured and presentation of domains per instrument

Instrument	Measurement of capability	Social wellbeing	Mental wellbeing	Physical wellbeing	Activity	Control	Other domains
ASCOT	Through wording, aims to measure an ideal state per domain	Social participation and involvement			Occupation	Control over daily life	Food and drink Personal cleanliness and comfort Personal safety Accommodation cleanliness and comfort Dignity
CALY	Currently unclear	Social relations	Health†	Health †	Occupation	Security	Time Financial situation Political resources Knowledge Living environment Housing
Capability-based questionnaire for assessing well-being in patients with chronic pain	Through wording of questions	Love and social inclusion Societal and family roles	Enjoyment Physical and mental wellbeing†	Physical and mental wellbeing†	Remaining physically and mentally active	Independence and autonomy	Respect and identity Feeling secure about the future
Child – and parent report questionnaire to explore capability of deaf children wearing a cochlear implant	First assess the level of functioning. Then adaptive follow-up questions inquiring about capability. See Appendix Table 2 for an example.	Relationship with parents Social participation Social skills Communication	Psychological well-being		School participation	Independence	Information access Assertiveness

Instrument	Measurement of capability	Social wellbeing	Mental wellbeing	Physical wellbeing	Activity	Control	Other domains
Diabetes specific instrument for measuring patient reported outcomes and experiences in the Swedish National Diabetes Register *	Through domains focusing on whether diabetes limits the person, and how individuals are able to deal with those limitations		How the patient feels What worries the patient	Barriers		Capabilities to care for your diabetes	Support from others Support from diabetes care provider Medical devices and medical treatment
ICECAP-A	Through wording of questions	Attachment	Enjoyment		Achievement	Stability	Autonomy
ICECAP-SCM	Through wording of questions	Love and affection Being supported	Emotional suffering	Physical suffering			Choice Dignity Preparation
ICECAP-O	Through wording of questions	Attachment	Enjoyment		Role	Control Security	

Instrument	Measurement of capability	Social wellbeing	Mental wellbeing	Physical wellbeing	Activity	Control	Other domains
Non-invasive prenatal testing related capability wellbeing	N.A. (instrument not developed at the time of writing this dissertation)	Affiliation	Emotions	Life Bodily health		Bodily integrity Control over one's environment	Senses, imagination and thought Practical reason Care taking (for existing or potential children and family)
OCAP-18	Through wording of questions, or asking about limitations in freedom	Affiliation	Emotions	Life Bodily health	Play	Control over one's life	Bodily integrity Senses, imagination and thought Practical reason Species
OxCAP-MH	Through wording of questions, or asking about limitations in capability	Affiliation	Emotions	Life Bodily health	Play	Control over one's environment	Bodily integrity Senses, imagination and thought Practical reason Species
Women's capability index	Through directly asking how much freedom someone has, wording of questions, and asking about limitations in freedom	Community relations Household wellbeing	Happiness Inner wellbeing	Physical strength			Economic security

Please note that the categorization of domains in different columns is supposed to give an overview of similarities and does not represent a complete comparison.
* Based on the domains included in the final instrument.
† One domain represents both physical and mental aspects of wellbeing.
Table published previously in Ubels et al. (2022b).

Despite the differences in the included domains, common elements across the instruments could still be identified. All of the instruments have content related to mental and social wellbeing (see Table 3). Furthermore, most of the instruments had content that covers physical health. It should however be noted that four instruments, the ICECAP-A, the ICECAP-O, the ASCOT, and the child – and parent report questionnaire to explore capability of deaf children wearing cochlear implants, do not have content that directly covers physical health (Al-Janabi et al., 2012; Grewal et al., 2006; Netten et al., 2012; Rijke et al., 2019).

2.3.4 Stage 3: The content of instruments and option freedom

Appendix Table 2 presents a comparison between the concept of option freedom and the content of the instruments for which items were developed at the time of writing this dissertation. More precisely, the content of these instruments was compared to the two components of the concept of option freedom: “options” and “access to options”. Some of the content of these instruments could not be classified according to these components. For example, some of the instruments contain items that seem more closely related to the experience of certain emotions, expressed in a way that is hard to classify as a freedom according to components of the concept of option freedom. These items were sorted into three additional categories: functionings, direct assessment of perceived freedom, measurement of general freedom (as opposed to individuals being able to access certain options), and other content which was identified by developers to be important for wellbeing assessment.

Generally, the instruments aim to assess capabilities through the wording of items and (Likert scale) response options. The wording of these items varied between the instruments, depending on how the developers of the instruments interpreted and applied the concept capability. Seven instruments aim to estimate the perceived capabilities of individuals in specific domains as objectively as possible (Al-Janabi et al., 2012; Engström et al., 2018; Greco et al., 2015; Grewal et al., 2006; Kinghorn et al., 2015; Lorgelly et al., 2015; Simon et al., 2013; Sutton and Coast, 2014). An example of such wording can be found in the ICECAP-A, which contains the item “I can have a lot of love, friendship and support” to measure an individual’s capability in the domain “attachment” (Al-Janabi et al., 2012). The developers of the ICECAP-A explicitly mention that their items were developed with the intention to measure capabilities as objectively as possible, instead of measuring a preferred level of

functioning (Al-Janabi et al., 2012). Other items and their response options were developed in such a way, that they more generally assess the perceived capability of an individual, instead of inquiring about a capability in a specific domain. An example can be found in the OCAP-18, an instrument that is based on Nussbaum's list of capabilities, which contains an item with the optimal response option "I am free to decide for myself how to live my life" (Lorgelly et al., 2008; Lorgelly et al., 2015).

Ten instruments have content that reflects the extent to which individuals are able to fulfill domains (Al-Janabi et al., 2012; Coast et al., 2008a; Engström et al., 2018; Greco, 2013; Kinghorn et al., 2015; Lorgelly et al., 2015; Netten et al., 2012; Rijke et al., 2019; Simon et al., 2013; Sutton and Coast, 2014). In relation to the concept of option freedom, these items reflect various options that are significant for individuals' wellbeing. The different response options of these items reflect whether individuals perceive to have the ability to realize those options. An example of such an item can be found in the ICECAP-A, where the domain "feeling settled and secure" has the optimal response option "I am able to feel settled and secure in all areas of my life" (Al-Janabi et al., 2012).

A selection of the instruments has content that does not focus on options but rather focuses on elements that influence the *access* to those options. Eight instruments have content that reflects if an individual experiences blocks or burdens in their access to realize certain options (Engström et al., 2018; Greco, 2013; Kinghorn et al., 2015; Lorgelly et al., 2015; Lorimer et al., 2007; Rijke et al., 2019; Simon et al., 2013; Sutton and Coast, 2014). For example, the item "Does your health in any way limit your daily activities, compared to most people of your age?" from the Oxford CAPabilities questionnaire-Mental Health (OxCAP-MH) covers whether health poses a block or a burden on the capability to do daily activities (Simon et al., 2013). Other instruments contain items that cover the level of support that individuals receive to access options, even though of the limitations that individuals might experience (Al-Janabi et al., 2012; Engström et al., 2018; Engström et al., 2016; Lorgelly et al., 2015; Netten et al., 2012; Rijke et al., 2019; Simon et al., 2013; Sutton and Coast, 2014). Particularly the capability based diabetes questionnaire and the ASCOT contain items that cover this type of content, with items such as "Do the support and services that you get from Social Services help you to maintain control over your daily life?" from the ASCOT (Engström et al., 2018; Engström et al., 2016; Netten et al., 2012).

Another observation was that developers included items in their instruments that could not directly be linked to the concept of capability. Instead of capabilities, these items covered (the absence of) various functionings that the developers considered to be important for an individual's wellbeing (Al-Janabi et al., 2012; Engström et al., 2016; Greco et al., 2015; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Lorgelly et al., 2015; Netten et al., 2012; Simon et al., 2013; Sutton and Coast, 2014). An example is the item "My health and mental well-being (including: pain, depression, sleep, mobility, medication side effects)" with the optimal response option "I have no problems with my physical health or mental wellbeing" (Kinghorn et al., 2015). Another type of functioning that is covered by a selection of instruments is the emotional experience of individuals (Engström et al., 2018; Greco, 2013; Kinghorn et al., 2015; Lorgelly et al., 2015; Netten et al., 2012; Simon et al., 2013; Sutton and Coast, 2014). For instance, the ICECAP-SCM contains the domain "emotional suffering", with the associated optimal response option "I rarely experience emotional suffering" (Sutton and Coast, 2014). Also, the instrument developed by Rijke et al. (2019) has content that does not directly reflect the capability concept. However, in the development of their instrument, which aims to assess wellbeing in children with cochlear implants, Rijke et al. (2019) clearly distinguished between capabilities and functionings in their a priori framework. Subsequently, there is a clear distinction between capabilities and functionings in the content of their instrument.

Other instruments contain domains and items, which seem to cover preferences instead of capabilities (Grewal et al., 2006; Kinghorn et al., 2015; Netten et al., 2012; Rijke et al., 2019; Sutton and Coast, 2014). These instruments assess whether individuals have achieved a preferred state in a certain domain. An example of a response option for such an item is "I have as much social contact as I want with people I like" from the ASCOT (Netten et al., 2012). In this case, developers interpreted the concept of capability as reflecting a preferred level of functioning.

The last observation was that there was a difference in the balance between two types of items across the instruments. The first type of item covers whether individuals have a certain level of capability or functioning in a certain domain. This type of item could also be described as "positively worded". The second type of item aims to measure the *lack* or *absence* of functionings or burdens in capability achievement which are detrimental to wellbeing. These items could be described as "negatively worded".

Some of the instruments contained only positively worded items. For example, the ICECAP-A consists of five items that cover whether individuals are able to fulfill capabilities in different domains (Al-Janabi et al., 2012). Other instruments contained a mix of positively and negatively worded items. For example, the ICECAP-SCM consists of five items that measure whether individuals are able to fulfill capabilities and two items that measure the presence of functionings that have a detrimental effect on wellbeing. The latter two items ask about the experience of physical and emotional suffering, with the optimal response options being the absence of physical or emotional problems. Negatively worded items are a noticeable part of instruments that were developed by conducting qualitative studies with participants that are affected by chronic disease (Engström et al., 2018; Kinghorn et al., 2015; Rijke et al., 2019).

2.4 Discussion

Generally, the developers of the instruments opted for using the capability approach as a framework due to its broad informational base in the context of wellbeing assessment. In this context, the developers of some of the instruments explicitly argued for the inclusion of a broad range of attributes that are important for wellbeing, while developers of other instruments stressed the need to assess wellbeing in terms of an individual's freedom. In order to develop an instrument, the researchers operationalized the measurement of capability in two steps. First, capabilities were identified that are important for the target population's wellbeing. Then, based on the identified capabilities, instruments were developed to assess wellbeing.

Most of the instruments are based on an a priori framework of capability. These frameworks are Nussbaum's list of capabilities or Sen's concept of capability. A priori concepts function as a "searchlight" that can highlight or obscure certain elements (Bordage, 2009). In the context of the capability approach, the conceptualization of capability by Sen and Nussbaum stresses that individuals need to have certain capabilities in order for them to be well-off. As argued, these conceptualizations focus on the "positive" elements of wellbeing. However, such positive a priori concepts of capability might hinder the identification of elements that block or burden capability achievement, which seems to be the case for some instruments. In the context of assessing capabilities, what is important is not only the capabilities themselves (or "options", from the concept of option freedom) but also how easily these capabilities can be realized (the "access to options") (Robeyns, 2017c). Capability instruments that

do not measure these blocks or burdens could therefore be limited in their ability to assess the capability wellbeing of individuals.

An additional observation was that in some instruments items could be identified which seem to be more related to functionings than capabilities, even though the developers of instruments stated the importance of assessing wellbeing in terms of capabilities due to its larger informational base. Of course, it is not necessarily problematic to assess wellbeing as a combination of functionings and capabilities. Indeed, assessing wellbeing as such might facilitate its assessment, given that functionings and capabilities represent different kinds of information. Sen himself called this combination of functioning and capability a “refined functioning” (Sen, 1985a, 1993b). A further discussion about refined functioning can be found in Fleurbaey (2006).

Still, instrument development would be facilitated if items measuring different concepts (e.g., functionings or capability) are differentiated from each other instead of being subsumed under the heading capability. Sen argued that subsuming or transferring two different concepts into a single entity might cause empirical problems in the application of those concepts (Sen, 1999b). Sen used the concept “utility” to illustrate this point. Utility has been understood in different ways, as has been explained in the introduction. In one understanding, utility represents an individual’s happiness. In another, utility is an expression of the observed choice behavior of an individual. Not differentiating between these two uses of utility, but merging both might lead to problematic conclusions, such as the notion that people behave in such a way, that their happiness is maximized. Of course, there are plenty of scenarios where people behave in a way that actively hurt their own chances of happiness (Sen, 1985b). In the context of the capability approach, subsuming different concepts under “capability” might result in difficulties for developers in deciding what kind of content to include in their instruments, since it is unclear which domains of wellbeing are best assessed in terms of capability and which domains are best assessed with alternative concepts, such as functioning.

To illustrate this point, Lorgelly et al. (2015) asked exactly this question to participants involved in the development of the OCAP-18. For example, for the domain “emotions”, participants were asked “What, in your view, do you value more? Being able to enjoy the love, care and support of your family and friends or actually enjoying it”. Lorgelly et

al. found differences between the domains of the OCAP-18. In some domains, such as “expressing views”, participants preferred to be evaluated in terms of capability. In other domains, participants showed mixed preferences, where some participants preferred to evaluate domains such as “love, care and support”, “planning of one’s own life” or “adequate nourishment” in terms of functionings and others in terms of capabilities. This exercise shows, that members of the public recognize the different types of information that capabilities and functionings represent and at times prefer one over the other in the context of wellbeing assessment. Given the different types of information that both concepts reflect, it is important for developers to make a clear distinction between these two concepts in their instruments, as is done by Rijke et al. (2019).

Further observations could be made regarding the content of capability instruments in relation to the concept of option freedom. The content of instruments that was developed with relatively disadvantaged participants, for example through chronic disease, cover more elements related to the blocks and burdens that people might experience whilst achieving capabilities. Furthermore, the developers of this type of instrument also included more content related to the measurement of functionings that have a detrimental effect on wellbeing. This effect of the inclusion of disadvantaged participants on the content of instruments has also been observed by Kinghorn (2015). While perhaps unsurprising, this observation has important implications for the development of instruments in the context of the capability approach, since it highlights that instruments that are developed with relatively advantaged participants might miss content that is important for evaluating capabilities in less advantaged individuals. For example, the ICECAP-A might be relatively insensitive to how physical health affects access to certain capabilities since it misses content covering this domain.

This is not only a theoretical discussion. Various empirical studies have been conducted regarding the sensitivity of conventional health economic instruments and newly developed capability instruments to different aspects of health. The results of these studies generally indicate that some of the capability instruments, such as the ICECAP-A, are relatively insensitive to physical health problems (Davis et al., 2013; Engel et al., 2017; Hackert et al., 2017; Khan and Richardson, 2018). In the context of HTA, this means that the informational base of some of the capability instruments is too limited to appropriately assess the effects of health technologies on their own.

A potential solution for these problems is by choosing a more comprehensive a priori concept of capability that incorporates the blocks and burdens that people might experience in their capability achievement. This choice can be guided by experts in the capability approach who are specialized in conceptual thinking. In this dissertation, the interpretation of the capability concept by Robeyns (2017c) is followed. The concept of option freedom is a more comprehensive understanding of capability compared to Sen's original definition which has been used up until now in instrument development. Besides facilitating the development of instruments to come to a completer assessment of capability, basing an instrument on a clearer concept also helps developers identify which domains are better evaluated in terms of capabilities, and which other domains might be better evaluated in terms of functionings or other elements that might potentially be important for wellbeing.

2.4.1 Limitations

A limitation of the review presented in this chapter is linked to the search strategy. It might be possible that capability instruments that are eligible for inclusion have not been identified, due to them not being referred to by the articles identified in the selection process. To reduce the chances of this happening, abstracts were screened by two researchers (all in the first round and 20% in the second round) and there was a broad selection of articles that were used as pearls in the first round. A second limitation is related to the search string that was used to identify pearls for the first search wave. These pearls were identified with the search string "capability approach". Some authors who work with and apply the capability approach use alternative terminology, for example, "capabilities approach". It is therefore possible that relevant publications and instruments are not included in this review.

Nevertheless, a majority of the instruments identified in this review explicitly used the capability approach as an a priori framework to develop an instrument. Furthermore, the developers of these instruments operationalized Sen's definition of capability in the development of their instruments. Given this observation, I believe that the findings in this review are relevant for researchers who are interested in developing instruments that are based on the capability approach.

2.4.2 Conclusion

Capability instruments have been developed to assess the wellbeing of individuals. This review shows that capability is generally understood as a form of freedom according to the conceptualization of Sen (Sen, 1993a). However, Sen's definition has been ambiguous (Robeyns, 2017c), which results in two problems in operationalizing the approach into instruments. The first issue is that some instruments are insensitive to some of the blocks and burdens that individuals experience whilst achieving capabilities. One reason for this problem is that due to the application of Sen's concept of capability, the developers of instruments did not include these blocks and burdens in their instruments. The second issue is that some of the content of the instruments do not reflect capabilities, but alternative concepts, such as functionings. This raises the question if some elements of wellbeing are better assessed in terms of functionings rather than capabilities.

In relation to HTA, these two observations imply that due to their content, some of the capability instruments might be relatively insensitive to certain effects of health technologies that are relevant for wellbeing, such as changes in experienced pain. Not measuring such effects might lead to suboptimal decision-making, given that policymakers that base their decisions on these instruments do not receive the necessary information to assess the effect of health technologies on wellbeing.

These issues can hypothetically be solved by developing instruments with a more comprehensive concept of capability. Developing instruments with a more comprehensive concept of capability would theoretically result in those instruments being able to more broadly assess an individual's capabilities. Furthermore, such a concept would facilitate researchers to differentiate between concepts (such as capability or functioning), which would result in items more accurately measuring different elements of wellbeing. In the context of HTA, such instruments more broadly inform the informational base that is used to assess the effect of health technologies. An example of such a more comprehensive concept is "option freedom" (Pettit, 2003). The next chapters will focus on operationalizing of the concept of option freedom into a capability instrument that can be used to assess the effect of health technologies on the wellbeing of individuals.

3 A THEORETICAL FRAMEWORK BASED ON OPTION FREEDOM

3.1 Background

The development of a conceptual framework is an important part of instrument development. Such a conceptual framework is arguably even more important in the context of the development of a capability instrument, given the normative implications of developing a capability list for evaluative purposes (see for a discussion Section 1.6.2).

As presented in Chapter 2, a range of capability instruments have been developed in a variety of different settings. Some instruments miss content related to the burdens that people experience, which is key for wellbeing assessment. These instruments lack this content due to a combination of using Sen's or Nussbaum's concepts of capability for instrument development and developing content with relatively healthy participants. Instruments developed with more disadvantaged individuals resulted in a completer assessment of capabilities that also included assessing the burdens that people experience in capability achievement. However, the content of these instruments is disease-specific and is not necessarily relevant for populations beyond the groups for which these instruments were developed.

A new framework for a capability instrument can however be developed by integrating the insights from the different development papers identified in the last chapter. Such a synthesis would need to be facilitated with a more comprehensive conceptualization of capability to identify all the relevant elements of capability wellbeing. Such a comprehensive conceptualization of capability can then also be used to identify whether elements of wellbeing are more appropriately assessed in terms of capabilities, functionings, or other constructs. The result of this synthesis would be a theoretical framework with a broad informational base that can be used to develop a new capability instrument.

One method that can be applied to conduct such a synthesis is the "best-fit framework synthesis" method (Carroll et al., 2013). The best-fit framework synthesis method has been described as a flexible and straightforward method of analyzing or synthesizing qualitative data (Carroll et al., 2013). In a best-fit framework synthesis, an a priori framework is identified and used to analyze qualitative data (Carroll et al., 2013).

During this analysis, the framework can be expanded or refined to better accommodate the qualitative data (Carroll et al., 2013). The result of a best-fit framework synthesis of qualitative studies is an updated model which encapsulates insights from each of these studies.

For the development of a capability instrument, such a synthesis has a further benefit. Authors have argued for researchers themselves to make an initial selection of capabilities and functionings that should be assessed (Coast et al., 2008b). This is related to the potential influence of problematic adaptive preferences from patients on their choice of relevant capabilities and functionings. Due to these adaptive preferences, potentially important capabilities and functionings for wellbeing assessment might be excluded. Consequently, a best-fit framework synthesis, as conducted in this study, might result in an initial list of capabilities and functionings for wellbeing assessment that can be validated with the involvement of the public at a later stage.

The aim of this chapter is to develop a conceptual framework based on the concept of option freedom, which can be used as the theoretical basis for an instrument that can be used to assess wellbeing.

3.2 Methods

The best-fit framework synthesis presented in this study follows three general steps. First, an a priori framework was identified. Then, relevant studies were identified that could be included in the synthesis. Lastly, the data from the studies were used for the qualitative analysis. In the current chapter, these qualitative data are derived from publications that support the development of instruments and the a priori framework is the concept of option freedom. New codes and themes are generated using thematic analysis techniques when data do not fit the a priori framework. The enhancing transparency in reporting the synthesis of qualitative research statement (ENTREQ) was followed to ensure that the results are presented transparently (Tong et al., 2012). The completed checklist can be found in Appendix Table 3.

3.2.1 The a priori framework

Due to the ambiguity of Sen's definition, which has been discussed earlier in this dissertation in Section 1.5.1, the decision was made to use an alternative

conceptualization of capability. It goes beyond the scope of this dissertation to give a complete review of the academic debate around the concept of freedom. Therefore, I followed expert opinion and use the concept of option freedom as an a priori framework of capability for the synthesis in this chapter (Pettit, 2003). “Options” represent the various opportunities that are available to individuals, and “access to options” represents how easy it is for individuals to realize these opportunities. Further information about the concept of option freedom can be found in Section 1.5.1.

3.2.2 Identification of relevant papers

The papers included in this study are a selection of the papers identified in Chapter 2. Development papers of capability instruments that contain “rich” qualitative data were included in the analysis. Development papers with rich qualitative data are those papers in which the qualitative work that is presented is supported with quotes from participants. The quality of the included development papers was evaluated with the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist, developed by Tong et al. (2007). Development papers of lower quality according to the COREQ checklist were however not excluded. Instead, the checklist was used as an aid, to ensure that I did not overlook any important indications of quality that could influence the development of the framework. Instead, a best-fit framework synthesis was conducted with all the identified papers and a post-hoc sensitivity analysis was conducted by excluding development papers with missing information (Carroll and Booth, 2015). In the post-hoc sensitivity analysis, the themes identified in the development papers with complete information according to the COREQ checklist were compared with development papers with missing information to establish whether certain parts of the theoretical framework were solely developed with papers with missing information. These parts would then be further scrutinized to ensure that they are still a meaningful component of the theoretical framework and not a consequence of misinterpretation of the qualitative data due to missing information.

3.2.3 Best-fit framework synthesis of data

Themes were developed with the aim that they could be operationalized as constructs in a capability instrument. This was done in three steps. In step (1), the complete results sections of the included development papers were extracted to Excel. In step (2), the data were deductively analyzed sentence by sentence with the a priori themes “options” and “access to options”, which were based on the concept of option freedom

(Pettit, 2003). In step (3), data that did not fit the a priori themes were then inductively analyzed with thematic synthesis methodology (Carroll et al., 2013). This resulted in new codes and themes. These emerged codes and themes were then re-applied to the data. This process was repeated to refine the themes and increase the coherence between the themes. The process stopped when no new themes emerged from the data and no further connections between the themes were made.

3.2.4 Information about the reviewers

The qualitative analysis was primarily conducted by me. I have a background in health sciences and health economics. To improve the quality of the analysis, the analysis itself and its results were discussed with two members of the Division of Health Economics of the German Cancer Research Center (DKFZ). EN is an anthropologist by training and is experienced in conducting qualitative studies. She provided valuable feedback on the coding process and the framework as it was being developed. The results were further discussed with KHV with a background in health economics. This resulted in valuable feedback to improve the clarity of the framework presented in this chapter.

3.3 Results

Of the 12 instruments identified in Chapter 2, 7 had associated published development papers that were eligible for inclusion in the current best-fit framework synthesis (Al-Janabi et al., 2012; Engström et al., 2018; Greco et al., 2015; Grewal et al., 2006; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Sutton and Coast, 2014). Publications related to the five other instruments were excluded since they did not have the rich data that are necessary for a best-fit framework synthesis (Lorgelly et al., 2015; Netten et al., 2012; Simon et al., 2013).

The assessment of the development papers according to the COREQ checklist resulted in several observations. Generally, the studies presented detailed information about how the content for the various instruments was generated. However, some development papers missed information according to the items of the COREQ checklist. Two development papers did not describe the relationship between the participants and the interviewers (Greco et al., 2015; Grewal et al., 2006). In four development papers, information about the personal characteristics of the interviewers was missing (Al-Janabi et al., 2012; Greco et al., 2015; Grewal et al., 2006; Kinghorn

et al., 2015). Consequently, it was difficult to evaluate how the personal characteristics of the interviewers and their relationship with the participants could have influenced the interpretation of qualitative data and the development of themes in these development papers.

3.3.1 Development of the framework

Four themes emerged from the best-fit framework synthesis: (1) “Option Wellbeing” (2) “Self-Realization” (3) “Perceived Access to Options” and (4) “Perceived Control”. The theme (1) “Option Wellbeing” represents that for individuals to experience wellbeing, they need to be able to access a range of options that result in life satisfaction. The theme (2) “Self-Realization” represents the importance of experiencing the ability to progress and develop oneself, and have a sense of meaning. The theme (3) “Perceived Access to Options” represents how individuals perceive their own ability to access options that are of value to them. This includes the perceived burdens that people experience whilst attempting to access those options. The theme (4) “Perceived Control” represents the individuals’ experience of being in control over their lives.

Each of these themes and associated subthemes will be discussed in detail in the next section. Figure 2 presents a graphical representation of the themes and their associated subthemes.

3.3.2 Option wellbeing

The best-fit frameworks synthesis started with two a priori themes: “options” and “access to options”. The theme “Option Wellbeing” is derived from the a priori concept “options” and represents the wellbeing derived from being able to exercise options. Various kinds of options are important for the experience of satisfaction for people. These options result in satisfaction when they can be exercised on what individuals perceive to be an adequate level. Not being able to access these options on a level that is experienced to be adequate harms individuals’ happiness and satisfaction. Several different abstract types of options were identified, which were categorized into five subthemes: “Emotional Wellbeing”, “Physical Wellbeing”, “Social Wellbeing”, “Environment Wellbeing” and “Activity Wellbeing”. It should be noted that these subthemes differ in their importance across development papers. This depends on the characteristics of the participants in the qualitative studies.

The subtheme “Emotional Wellbeing” represents the importance of being able to be happy, have pleasure, and be content. Conversely, it also covers the ability to not be affected by sadness or feelings of depression. All the development papers noted this to be an important element of wellbeing (Al-Janabi et al., 2012; Engström et al., 2018; Greco et al., 2015; Grewal et al., 2006; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Sutton and Coast, 2014).

> *“…obviously it’s [mother’s illness] been hard, it’s been upsetting…and visiting her now isn’t exactly a barrelful of laughs… I guess it’s saddening …” (Female, 29), Al-Janabi et al. (2012).*

The subtheme “Physical Wellbeing” represents the importance of being physically healthy. This also includes the absence of pain or discomfort. The subtheme “Physical Wellbeing” is prominently described in development papers where qualitative research was conducted with participants affected by (chronic) disease (Engström et al., 2016; Kinghorn et al., 2015; Sutton and Coast, 2014), in development papers with elderly participants (Grewal et al., 2006), as well as in development papers that aim to assess the impact of new health technologies with potential users (Kibel and Vanstone, 2017).

> *“Pain is very tiring” (Female, not employed), Kinghorn et al. (2015).*

The subtheme “Social Wellbeing” represents the wellbeing that is derived from social contact with friends or family. Its importance was noted in all the development papers that were analyzed in this chapter (Al-Janabi et al., 2012; Engström et al., 2018; Greco et al., 2015; Grewal et al., 2006; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Sutton and Coast, 2014).

> *“At ante-natal classes …six of us really gelled and just became the closest of friends. It was like we’d known each other for years and years and years. … we see each other all of the time and we help each other out which is great.” (Female, 32), Al-Janabi et al. (2012).*

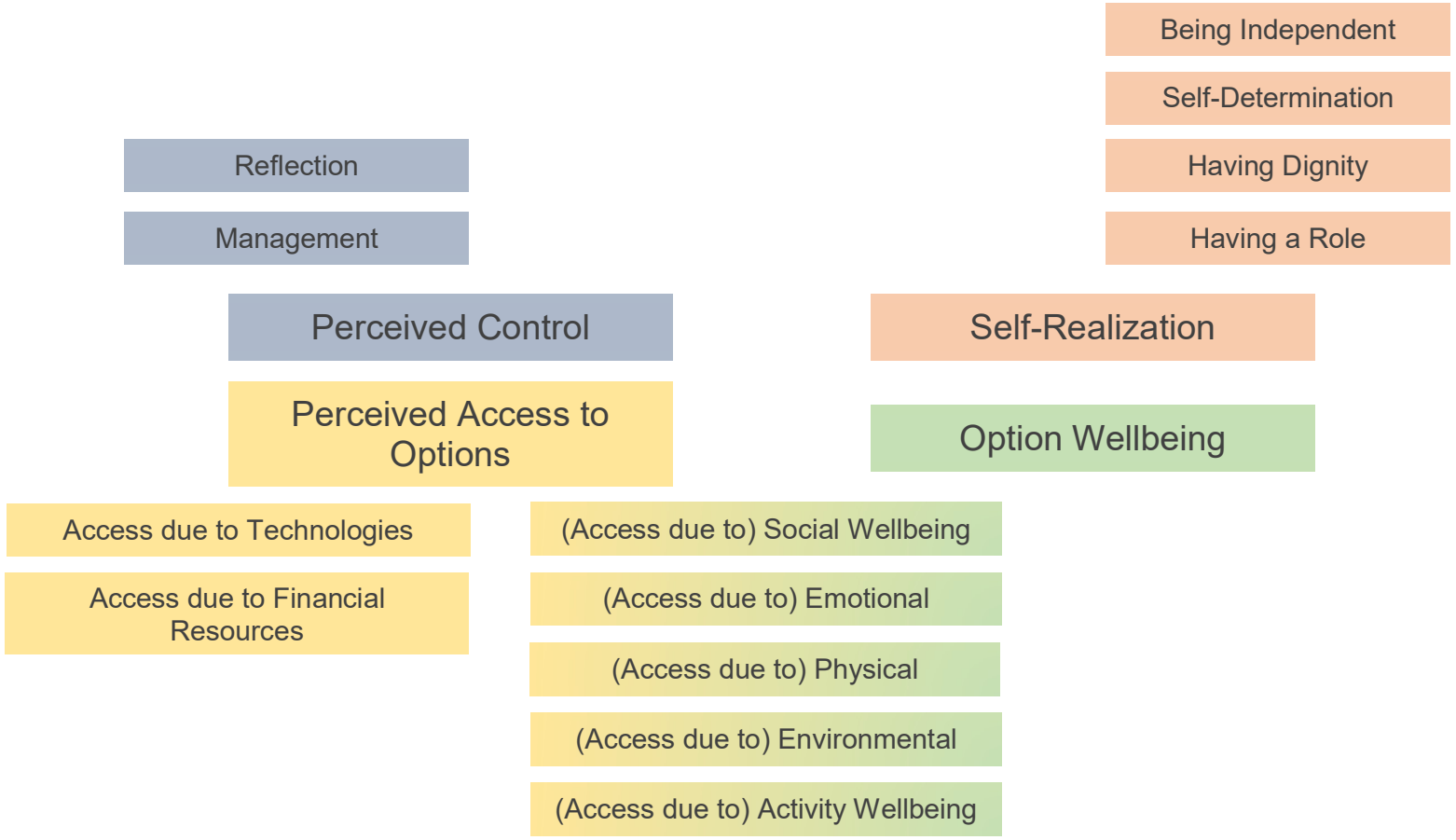


Figure 2. Graphical representation of themes and subthemes.

Figure has previously been published in Ubels et al. (2022a).

The subtheme “Environment Wellbeing” stresses the importance of being settled in their immediate environment (such as their home or neighborhood) for people to experience wellbeing. Its importance emerged in the analysis of four development papers (Al-Janabi et al., 2012; Greco et al., 2015; Grewal et al., 2006; Kinghorn et al., 2015).

> *“I suppose it’s security, ain’t it—you’ve been in the area for a long time, you know where everything is, you know everybody locally—you feel more secure...” (Male, aged 69), Sutton and Coast (2014).*

The subtheme “Activity Wellbeing” represents the importance of being able to do things for fun or relaxation. This also has a positive impact on wellbeing. This subtheme also emerged from the analysis of four development papers (Al-Janabi et al., 2012; Greco et al., 2015; Grewal et al., 2006; Kinghorn et al., 2015).

> *“I can spend all day long in art galleries—I love art” (Female, aged 85), Sutton and Coast (2014).*

3.3.3 Self-realization

The theme “Self-Realization” emerged from the observation that elements were mentioned in the development papers that are different from the experience of satisfaction but are still important for wellbeing. These elements were categorized into four subthemes: “Having a Role”, “Having Dignity”, “Being Independent” and “Self-Determination”.

The subtheme “Having a Role” represents the observation that there are certain activities that are important for wellbeing beyond generating happiness or satisfaction. These activities give a sense of meaning, identity, or self-worth (Al-Janabi et al., 2012; Greco et al., 2015; Grewal et al., 2006; Kinghorn et al., 2015). For instance, Kinghorn et al. (2015) observed that due to chronic pain, some men felt less masculine, since they could not carry heavy things to support their partners.

> *“…you feel inadequate. Well I do, when my missus starts…unloading the car, and I walk into the house and sit down.” (Male employed), Kinghorn et al. (2015).*

The subtheme “Having Dignity” represents the level of respect that individuals perceive to receive in society. Furthermore, it represents the perceived ability of individuals to be recognized as beings of worth. The importance of being recognized was noted by several development papers (Al-Janabi et al., 2012; Engström et al., 2016; Greco et al., 2015; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Sutton and Coast, 2014). Furthermore, in two development papers, the importance of being able to behave in a way that gives a sense of self-respect is noted in two development papers (Greco et al., 2015; Sutton and Coast, 2014).

> *“I’ve got my self-respect, she [caretaker] doesn’t stand there if I’m having a shower and all that, she just makes sure the windows are covered … we all want our self-respect no matter who we are.” (Female, 68 years), Sutton and Coast (2014).*

The subtheme “Self-Determination” represents the importance for an individual of being able to take direction over one’s life and being able to choose options that are valuable to an individual. This subtheme was reflected with different kinds of wording in the development papers. For example, Greco et al. (2015) documented that individuals find it important to express themselves without being oppressed. Furthermore, Greco et al. (2015) noted the importance of the ability to do certain activities without the need to ask for permission. Al-Janabi et al. (2012) described the importance being able to achieve goals and move forwards in life for individuals’ wellbeing. Kibel and Vanstone (2017) and Sutton and Coast (2014) reported that individuals stress the importance of the ability to make choices about elements that influence their lives.

> *“I’m staying here until I get carried away. I’ve worked hard and paid for it, and this is my abode and I’m quite happy with it.” (Male, 72 years) Sutton and Coast (2014).*

3.3.4 Perceived access to options

The theme “Perceived Access to Options” is a direct reflection of the a priori theme “access to options”. The theme represents the individuals’ perceived ability to access options that are important to them, and the burdens that they might experience while doing so. The reason for adding the word “perceived” was to stress the subjective

nature of this access, given that it is developed with development papers that contain information from individuals' experiences.

The subthemes associated with the theme "Perceived Access to Options" are linked to the subthemes under the theme "Option Wellbeing". Lower levels of "Physical Wellbeing" were reported to burden or block access to options (Al-Janabi et al., 2012; Engström et al., 2016; Greco et al., 2015; Grewal et al., 2006; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Sutton and Coast, 2014).

> *"The negative thing with diabetes is when people at work ask if you can join them for something after work. No, I can't, I'm going home to take my injection and have dinner. You get a little tied up, you know." (Female, 60 years old, Type 2 Diabetes Mellitus), Engström et al. (2016).*

Similarly, several development papers report the effect of "Emotional Wellbeing" on individuals' ability to access options (Engström et al., 2018; Greco, 2013; Kinghorn et al., 2015). For example, a quote from one participant of one of the studies illustrates that being emotionally well-off makes it easier to access other options.

> *"If you feel mentally good and feel that everything is going well, then you have more energy to take care of your diabetes. You, like, want to feel good physically too." (#29; Female, 22 years old, Type 1 Diabetes Mellitus), Engström et al. (2016).*

Researchers of several development papers also noted the effect of "Social Wellbeing" on the way that individuals can access options, with higher satisfaction in this subtheme facilitating the ability to access options (Al-Janabi et al., 2012; Engström et al., 2016; Greco et al., 2015; Grewal et al., 2006; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Sutton and Coast, 2014).

> *"When people cooperate [in the household] it becomes easy to develop [prosper]. And you have a good life. You live peacefully in the home." Greco et al. (2015).*

Aspects related to “Environmental Wellbeing” were also mentioned in several development papers as a factor that influences the ability of individuals to access options (Al-Janabi et al., 2012; Greco et al., 2015; Grewal et al., 2006; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Sutton and Coast, 2014).

> *“It was lovely down here at one time, but it’s just frightening now… I wouldn’t walk when it were dusk from here to the top of the hill… because there’s weirdos on the canal.” (Female, aged 66), Grewal et al. (2006).*

Two further subthemes emerged that were not associated with the subthemes under “Option Wellbeing”, but only relate to individuals’ capabilities. These subthemes are called “Access due to Financial Resources” and “Access due to Technologies”. They can be seen as means to reach subjective wellbeing. The subtheme “Access due to Financial Resources” represents how well-off individuals perceive to be in terms of their finances. In several development papers, it is noted how financial resources facilitate accessing various options (Al-Janabi et al., 2012; Engström et al., 2016; Greco, 2013; Grewal et al., 2006; Kibel and Vanstone, 2017; Kinghorn et al., 2015).

> *“I’m reasonably fortunate… in so far as that we’ve got the two pensions… we’re able to go off... We grabbed a cheapie flight at the end of April... flew down to Nice...” (male, aged 70), Grewal et al. (2006).*

The second subtheme is “Access due to Technologies”. The use of medical and non-medical technologies mitigated the impact of blocks and burdens in access caused by health problems, such as devices helping individuals affected by diabetes to manage their symptoms (Engström et al., 2016). Furthermore, technologies also created the opportunity to access new options, such as non-invasive prenatal testing to screen for genetic abnormalities (Kibel and Vanstone, 2017). The important role that these types of technologies play in individuals’ lives was noted in several development papers (Engström et al., 2016; Greco, 2013; Grewal et al., 2006; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Sutton and Coast, 2014).

> *“I think the insulin pump is fantastic. Because it gives me freedom.” (#24; Female, 64 years old, Type 1 Diabetes Mellitus), Engström et al. (2016).*

Lastly, it should be noted throughout the quotes presented above, individuals reflect on the ease of access to options in light of the options that are important to them. To illustrate, the quote from the 66-year-old woman in the study by Grewal et al. (2006) in the subtheme "Access due to Environmental Wellbeing" illustrates how she experiences a reduction in freedom because of being unable to go for a walk in the evening. Consequentially, the perceived ease of access represents both access to options itself, as well as the availability of options with importance to an individual. As such, this theme represents the perceived capabilities of individuals.

3.3.5 Perceived control

The theme "Perceived Control" represents the idea that it is important for people to "hold the reins" over their own lives. This experience is influenced by a balancing act between the options that an individual is able to access and the options that are actually valuable for individuals, but might not be accessible due to experienced blocks or burdens. Individuals' experienced control over their lives is a consequence of managing those blocks and burdens in such a way, that individuals are able to access those options that are valuable to them. In several development papers, authors have noted the importance of this perception of control in relation to wellbeing (Al-Janabi et al., 2012; Greco et al., 2015; Grewal et al., 2006; Kibel and Vanstone, 2017; Kinghorn et al., 2015). Linked to this theme are the subthemes "Management" and "Evaluation".

It is important to note that the subthemes related to the theme "Perceived Access to Options" are mutually interdependent. In some cases, limitations in access due to a lower level of achievement in one option can be compensated with higher levels of achievement in other options. The subtheme "Management" represents the strategic nature of dealing with these limitations in access to options. For example, a quote from a participant in the development paper of Grewal et al. (2006) shows that options that are limited due to a lack of financial resources can be compensated with having higher levels of social wellbeing.

> *"… you just see the predicament, you don't see anything else, but a terrible disaster… and I thought: 'Whatever am I gonna do?' and that night he just stood up in front and he said: 'One of our ladies has hit a problem'. And I thought: 'Oh no, you're not'. And he did. He said: 'I'd like you to feel that you like to help*

(informant) in her predicament', and you know, £300 went in that basket and I cried and I cried and I cried." (Female, aged 71), Grewal et al. (2006).

At times, dealing with burdens or blocks in access to one set of options resulted in the access to other options being blocked or burdened. For example, one participant reported how reducing the burden of pain through taking painkillers resulted in an additional burden due to the need to plan when to take medication.

"This morning, I got up—5 o'clock—I took my first pain killers, went back to bed again so that I was ready to get up to have my shower at half past six, or else, by the time you start taking them they haven't taken effect and you're trying to move around. So, yeah, you've got to think ahead…" (Female, not employed), Kinghorn et al. (2015).

The strategic nature of "Management" is also reflected in how individuals evaluate and pursue those kinds of options that are most important to them, even if that comes at the cost of not being able to access other options (Engström et al., 2016; Greco et al., 2015; Kibel and Vanstone, 2017; Kinghorn et al., 2015; Sutton and Coast, 2014).

"I would be so pleased if I didn't have to pay for the CGM [continuous glucose monitoring] myself (...) If I couldn't afford to pay for it, I wouldn't have such a good blood count [Hemoglobin A1c]" (Female, type 1 Diabetes Mellitus, 44 years old), Engström et al. (2016).

The subtheme "Evaluation" is related to an individual's perception of control on a more fundamental level. The subtheme represents an individual's evaluation of their actual ability to access options and their preferred level of access to options. As such, the blocks and burdens in option access influence this evaluation. This evaluation is also influenced by the individual's expectations regarding how burdens and blocks develop over time (Al-Janabi et al., 2012; Engström et al., 2016; Grewal et al., 2006; Kinghorn et al., 2015; Sutton and Coast, 2014). Adaptation of preferred option levels might happen if the blocks or burdens on access are high.

"That would be a good point to put next to the hobbies, er walking…. I'm not able to do that, so then I had to look for something else to occupy my mind. So right,

then—I always did do a lot of knitting and cross stitch—I take a lot of interest in that, and reading." (Female, Retired), Kinghorn et al. (2015).

Some authors noted that acceptance is an important way for individuals to adapt to blocks or burdens in access to options (Engström et al., 2016; Kinghorn et al., 2015).

3.3.6 Relationships amongst themes

The relationship between the themes is complex. However, in general terms, burdens and blocks in access to options seemed to harm subjective wellbeing when those burdens and blocks are challenging to deal with and preferences are not adapted (Engström et al., 2016; Greco et al., 2015; Kinghorn et al., 2015). This negative impact particularly translated itself into lower levels of "Emotional Wellbeing" (Al-Janabi et al., 2012; Grewal et al., 2006; Kibel and Vanstone, 2017) and lower levels of "Perceived Control" (Engström et al., 2016; Grewal et al., 2006; Kinghorn et al., 2015).

"…my health broke down again … which came as a shock… I had to give up work immediately …and it cast a long shadow because it's always there in the background, you never know when it might jump on you. So you live with uncertainty." (Female, 78), Al-Janabi et al. (2012)

When individuals did manage to live with blocks and burdens, Engström et al. (2016) reported that individuals could experience a sense of pride.

"Being able to manage diabetes in different situations could be related to feeling proud and a sense of trusting one's own ability." (authors' observation), Engström et al. (2016).

This indicates that effectively managing blocks and burdens could increase the experienced wellbeing of individuals to a certain extent.

3.3.7 Post-hoc sensitivity analysis

As was presented earlier in the results, some information was missing in four development papers in the context of the COREQ checklist (Al-Janabi et al., 2012; Greco et al., 2015; Grewal et al., 2006; Kinghorn et al., 2015). These development papers were excluded in a post-hoc analysis of the development papers. One of the

results of the post-hoc sensitivity analysis was, that all the themes and their related subthemes could be identified in the three remaining papers (Engström et al., 2018; Kibel and Vanstone, 2017; Sutton and Coast, 2014). This indicates that the emerged themes and subthemes were not too affected by the suboptimal presentation of qualitative research, which speaks for the strength of the developed framework.

3.4 Discussion

The aim of this chapter was to develop a theoretical framework that can be used for the development of a capability instrument. In light of the guidelines for instrument development presented in Section 1.6.1, this framework should be understood as an important first step, since it describes the constructs that should be measured to assess health-related capability wellbeing. The framework has been developed with an a priori concept of freedom to ensure that relevant elements related to the assessment of capability are identified. The concept of option freedom proved to be a useful a priori framework to conduct the synthesis (Pettit, 2003). As an a priori framework, the concept helped to identify the differences between options, the ways in which these options can be accessed, and how different elements can influence the accessibility of options. Furthermore, the concept of option freedom also helped to distinguish between capabilities and functionings related to the subjective experience of those capabilities, such as happiness or social wellbeing. Indeed, one result of the analysis was that both capabilities and functionings give unique information regarding the wellbeing of individuals.

A result of the synthesis is that the a priori themes “options” and “access to options” were adjusted to reflect the experiences of individuals. The a priori concept “access to options” was changed into the theme “Perceived Access to Options”. Through this change, I intend to signify the importance of an individual’s perception regarding the accessibility of options. The a priori concept “options” changed into two different themes that reflect two types of functionings: “Option Wellbeing” and “Self-Realization”. The theme “Option Wellbeing” represents the satisfaction derived from having options. The theme “Self-Realization” reflects whether individuals can experience having a meaningful life with the options available to them. The theme “Perceived Control” emerged through inductive analysis of the qualitative data.

The subthemes that are linked to both the themes “Option Wellbeing” and “Self-Realization” show similarities to elements that are associated with wellbeing in other fields of research (Alkire, 2002; Qizilbash, 2002). For example, the subthemes “Social Wellbeing”, “Physical Wellbeing”, and “Having a Role” reflect elements of wellbeing that are shared in different disciplines (Alkire, 2002). This is perhaps unsurprising. Several authors, such as Alkire (2002) and Qizilbash (2002), have argued that such lists of elements that are important for wellbeing share commonalities across research traditions. Indeed, Qizilbash (2002) argues that these lists are based on common grounds and that the variety between lists rather reflects differences in research context than fundamentally different values. The subthemes linked to the themes “Option Wellbeing” and “Self-Realization” thus seem to mirror elements of wellbeing that are commonly mentioned in literature.

As a theme, “Perceived Access to Options” represents the individuals’ perceived ability to realize or achieve options. This perceived ability is affected by the burdens that individuals experience, which might hinder but not block their access to options. Still, such burdens harm individuals’ capabilities. Particularly in the field of health, the assessment of these burdens might be instrumental to understand the capabilities of individuals. For example, an individual with chronic pain might be able to follow classes, just like an individual without chronic pain. From the perspective of available options, there is thus no difference between these two individuals. However, once the extra effort is taken into account that an individual with chronic pain might have to exert to control the pain to follow these classes, it could still be argued that the individual without chronic pain has more capabilities than the individual with chronic pain. The assessment of these burdens is thus key for the evaluation of capabilities (Robeyns, 2017a).

Additionally, the “Perceived Access to Options” theme represents the capability concept in the theoretical framework developed in this chapter. As was explained in Section 1.5.1 of this dissertation, a capability is best understood as consisting of options and access to those options. As was highlighted in this chapter, individuals assess their own capabilities in terms of the ease of access to options that are important to them. When operationalizing the concept of option freedom into a self-report instrument, it might therefore be sufficient to measure an individual’s perceived ease to access those options to assess their capabilities.

One caveat in this context is adapted preferences. What if the importance of options to individuals changes due to them being impaired? Could this lead to individuals adapting and considering other options to be more important, which results in individuals experiencing a high level of capability even though in practice, they might have lost some options? To illustrate, an immobile individual who considers reading to be an important option might report having easy access to options as long as she can read. The fact that she is not mobile is thus not reflected in her considerations of her capabilities. When operationalizing the framework of this chapter into an instrument, this could be problematic, since adaptation of preferences could affect the responses to self-report instruments inquiring about the perceived access to options of an individual. A potential method of testing whether such problems exist will be discussed in Chapter 5 of this dissertation.

One theme emerged from the inductive analysis that did not fit the a priori themes: “Perceived Control”. The theme reflects the positive influence of having the experience to be in control of individuals’ experienced wellbeing. Based on my interpretation of the qualitative data, the experience of control might even mitigate the effect of burdens or limitations in access to options and experience to be well-off. In this respect, a parallel can be drawn between this theme and capability literature. Bellanca et al. (2011) introduced the concept “dis-capability”. They argue, that an individual can be considered dis-capable when she or he is unable to manage and/or adapt to limitations in capability. Individuals who can adapt and manage these limitations can thus be considered well-off according to Bellanca et al. (2011). This was also reflected in the development papers, where participants mentioned to consider themselves well-off once they felt they were in control over their disadvantage.

Two subthemes were linked to the theme “Perceived Control”: “Management” and “Evaluation”. The subtheme “Management” represents the different strategies that individuals use to deal with limitations and burdens in access. The type of strategy that individuals use depends on the availability of resources and how the management of one limitation affects the accessibility to options. Controlling limitations thus requires individuals to make strategic choices. These choices have also been observed in other empirical studies. Gibbins et al. (2014) studied how advanced cancer patients managed their chronic pain. They found that patients consumed their medication to

control the limitations that are a consequence of pain, but to a level that minimized the limitations imposed by side effects of the medication. Furthermore, Gibbins et al. (2014) found that regarding the subjective wellbeing of patients, what actually mattered more was the effect of pain and the side effects of medication on patients' independence than the actual experience of pain itself. Similar observations regarding the importance of patients having control over the limitations that are caused by various types of disease are found in other studies as well (Albrecht and Devlieger, 1999; Connell et al., 2012; Vallerand et al., 2007; Zeppetella, 1999).

The subtheme "Evaluation" represents that individuals evaluate which options are important to them, in light of the options that they prefer and the available options. Individuals experience a sense of control when there is an alignment between the options that they prefer and the available options. However, if the difference between preferred options and available options is too large, a process of revaluation happens which might result in recognizing the value of alternative options. This subtheme shows parallels with the concept "adapted preferences", which was introduced in Section 1.4 of this dissertation. Both concepts represent how people change their preferences in light of what is possible for them. In the field of health, facilitating the adaption of individuals to their chronic diseases when there are no other possibilities to reduce their effects has also been seen as an integral part of treatment, given that it is the only option for these individuals to experience wellbeing (Sprangers and Schwartz, 1999).

3.4.1 Limitations

There are several limitations associated with the synthesis presented in this chapter. First of all, the framework that has been developed is based on the synthesis of qualitative data from seven papers. Additional qualitative data could have improved the trustworthiness of the framework (also called credibility, see Lincoln and Guba (1985)). Furthermore, the coding of the data was largely conducted by me. Researcher triangulation took the form of discussing the framework and explaining the reasoning behind the coding with experts, however, it did not comprise of two researchers independently coding the same data. This as well reduces the credibility of the analysis (Lincoln and Guba, 1985).

3.4.2 Conclusion

To conclude, this chapter presents the first step of instrument development by presenting the development of a framework of constructs that will need to be measured to assess health-related capability wellbeing. This framework consists of four themes. Compared to Sen's concept of capability, this framework has two advantages. First, it acknowledges that individuals who can access options with ease cannot be considered to have a similar level of wellbeing to individuals who can access the same options with difficulty. Second, the informational base of wellbeing in the framework does not only concern capabilities but also functionings related to subjective wellbeing and the experience of control. The next step in instrument development is to establish whether the themes can be operationalized as constructs in an instrument. This will be the subject of the next chapter.

4 OPERATIONALIZING THE FRAMEWORK INTO AN INSTRUMENT

4.1 Introduction

In the last chapter, I presented the development of a theoretical framework for wellbeing assessment that is based on the concept of option freedom. Before applying this list in a survey study, it has to be established that these themes can be operationalized as constructs in an instrument. This step is important since establishing what types of constructs are measured helps researchers understand how individuals interpret and respond to items (a more extensive explanation about why this is important can be found in Section 1.6). Furthermore, first evidence can be developed for the feasibility of measuring capabilities in terms of “Perceived Access to Options”. Lastly, an instrument could function as a proof of concept that shows how a capability instrument based on the theoretical framework of the last chapter differs from existing capability instruments.

With respect to HTA, it is also important to know which constructs are being measured, since it informs researchers about what aspects of wellbeing an instrument is sensitive to, which affects an instrument’s ability to measure changes in wellbeing. This is important information since depending on the constructs included in an instrument, the measured effects of a health technology might differ dramatically (Khan and Richardson, 2018).

The translation of themes to constructs, as well as studying the content of these constructs themselves is thus a key step in the development of a health-related capability instrument. Therefore, the overall aim of this chapter is to operationalize the theoretical framework from Chapter 3 into such an instrument. In this chapter, this is done through three sub-aims, which are loosely based on the guidelines for instrument development by Boateng et al. (2018) that are presented in Section 1.6.1. The first sub-aim is to identify relevant items that could potentially be used to measure the themes developed in the theoretical framework from the last chapter. The second sub-aim is to conduct a confirmatory factor analysis to study whether the themes can be translated into constructs in a confirmatory factor analysis. The third sub-aim is to develop an instrument based on insights from the confirmatory factor analysis.

4.2 Methods

4.2.1 Data

The data that were used for the analysis in this chapter comes from the Multi-Instrument Comparison (MIC) study (Richardson et al., 2012). The MIC study aims to study and compare different HRQoL and wellbeing instruments. The core questionnaire which was administered to every participant of the MIC study consisted of twelve instruments, with additional items related to the demographics of participants. Out of these twelve instruments, three are used to assess subjective wellbeing: the Personal Wellbeing Index (PWI) by Cummins et al. (2003), the Integrated Household Survey from the Office of National Statistics (ONS) by Dolan and Metcalfe (2012) and the Satisfaction with Life Survey (SWLS) developed by Diener et al. (1985). Six of these were multi-attribute utility instruments: the EuroQol five-dimensional instrument with five response options (EQ-5D-5L) by Herdman et al. (2011), the Assessment of Quality of Life instrument (AQoL) with eight dimensions (AQoL-8D) developed by Richardson et al. (2009), and with four dimensions (AQoL-4D) by Hawthorne et al. (1999), the Fifteen Dimensional measure of health-related quality of life (15D) by Sintonen and Pekurinen (1993) and the Quality of Well Being Self-Administered (QWB-SA) scale (Kaplan et al., 1993; Seiber et al., 2008). Additionally, three non-utility instruments were included in the MIC survey: the 36-Item Short Form Health Survey Version 2 (SF-36 V2) by Brazier et al. (2002), the ICEpop CAPability measure for Adults (ICECAP-A) by Al-Janabi et al. (2012), and a TTO exercise (see Section 1.3.2 for an explanation about TTO). Combined, this resulted in a questionnaire consisting of 227 items that was distributed to the complete sample. Further disease specific instruments were additionally administered to the main questionnaire to certain groups of participants. The items in the MIC study were generally constructed as Likert scales (DeVellis, 2017b), with item response options ranging from two to eleven response options.

The MIC study followed a cross-sectional design and was conducted in six countries (Australia, Canada, Germany, Norway, United Kingdom, and the USA). Participants of the MIC study were recruited through a survey company. Participants in their database were asked to complete an online version of the questionnaire. These participants were recruited to fill quotas, such that enough participants with different health conditions were included. One group consisted of individuals who were considered

healthy. Nine other groups consisted of individuals that were affected by different diseases (arthritis, asthma, cancer, depression, diabetes, hearing loss, heart problems, stroke, and chronic obstructive pulmonary disease). The researchers who constructed the dataset performed several edits to remove anomalous responses (Richardson et al., 2012). These are presented in Appendix Section 10.3.1.

To avoid priming the participants' responses, the MIC survey started with the subjective wellbeing instruments. These instruments were administered in their entirety but were randomized in terms of their order. This was followed by a question about whether the participants had a certain disease, followed up by a confirmatory question regarding their disease status. Then, participants were asked to complete the demographic items. Participants affected by diseases were then asked to complete a disease-specific questionnaire. The MIC survey concluded with HRQoL instruments. The order in which these instruments were presented was randomized. Further information about the MIC study and its survey can be found in Richardson et al. (2012) and the website of the MIC project (https://www.aqol.com.au/index.php/aqol-current).

For the analysis presented in this chapter, the complete dataset was randomly divided into two subsets. One subset functioned as a "training" dataset (n = 4011) and was used to initially test and refine the measurement model. The second "test" dataset was used to validate the measurement model after refining it (n = 4011). The items within the database were recoded, with higher scores meaning that individuals are better off (e.g., individuals with a high score on a certain item experience less pain, are happier, and are less anxious).

4.2.2 Selection of items

A selection process was conducted to select items for the construction of a measurement model. First, to increase the measurement precision, items were excluded from the selection when they had fewer than four response options (Simms et al., 2019). Then, the framework developed in Chapter 3 formed the basis for identifying relevant items in the MIC database for the measurement model. The themes, subthemes, and their associated quotes were compared with the wording of items in the MIC database to develop an item bank. Items were included in the item bank when there were similarities between the wording of an item and a quote. These items were used to assess the construct that is associated with the quotes.

From the full MIC survey questionnaire, 56 items were selected for the confirmatory factor analysis. These 56 items came from seven instruments. Three of these were HRQoL instruments: the SF-36 V2 (Brazier et al. (2002), the AQoL-4D, (Hawthorne et al., 1999), and the AQoL-8D (Richardson et al., 2009). Three of the seven instruments are used to assess subjective wellbeing: the SWLS (Diener et al., 1985), the ONS (Dolan and Metcalfe, 2012), and the PWI (Cummins et al., 2003). Also, items from a capability instrument were included, the ICECAP-A (Al-Janabi et al., 2012). Appendix Table 4 gives an overview of the selected items and associated instruments. These items were recoded in such a way that a higher score means that individuals have fewer problems (for example less pain) or a higher level of wellbeing.

4.2.3 Analysis methods

A confirmatory factor analysis (see Section 1.6.1 for an explanation of confirmatory factor analysis) was conducted in R version 4.3.0 (R Core Team, 2013). The analysis was conducted with the “Lavaan” package version 0.6-15 (Rosseel, 2012). Further psychometric testing of the scales was conducted with the “Psych” package (Revelle and Revelle, 2015).

A measurement model was developed with items from the MIC study database. The measurement model was tested in two stages. In the first stage, a theory-driven confirmatory factor analysis was conducted to study whether the themes from the framework developed in Chapter 3 could be measured as factors in a measurement model. These factors consisted of the items that were selected by linking the qualitative framework with items from the MIC database (see the selected items and their link to the qualitative framework Appendix Table 4). The measurement model was then further modified with theoretically guided adjustments to see if the fit of the measurement model could be improved. It should be noted that throughout the development of the measurement model, the main constructs of interest (initially “Option Wellbeing”, “Self-Realization”, “Perceived Access to Options” and “Perceived Control”) were allowed to correlate with each other.

In the second stage, the measurement model was further adjusted with data-driven improvements. These adjustments were based on the assumption that some of the

model misfit could be explained by the presentation of the items in instruments (for example testlets) or by unaccounted similarities between different items (for example, multiple items inquiring about experienced pain). Residual correlations higher than 0.1 and the highest values of modification indices were used to identify sources of misfit. The MIC questionnaire was studied after the identification of a source of misfit. Special attention was given to the content of those items that showed misfit. For example, multiple items were identified that were related to the experience of happiness. Furthermore, the potential effect of the layout of the MIC questionnaire was studied, which could have been a source of common method variance. Adjustments to the model were made when the content of items showed similarities or the layout of the questionnaire of the MIC questionnaire might have been a cause of these sources of misfit.

Several strategies were used to account for misfit. Strategy (1) was to construct orthogonal factors to account for the residual correlations amongst items with overlapping content. Strategy (2) was to construct orthogonal factors that account for correlations that could be attributed to common methods, such as testlets or response option length (MacKenzie and Podsakoff, 2012). For the construction of these specific orthogonal factors, a minimum of three items had to show residual correlations higher than 0.1 with each other. The use of this type of specific factors is common in the development of bifactor models (Reise, 2012). Item correlations in bifactor models are modeled by a common factor and multiple specific (also called group) factors (Rodriguez et al., 2016). The common factor accounts for the shared variance of all the items in a dataset. The specific factor accounts for the variance of items that share similar content.

At times, only pairs of items showed strong correlations with each other. For these items, it was not possible to develop separate orthogonal factors. In this case, strategy (3) was to correlate the error terms of these items. Another source of misfit was that items showed residual correlations with other items outside their own factor. To account for these correlations, strategy (4) was to let these items cross-load on different factors when there was an overlap between the content of the item and the factor represented overlapped.

One key assumption in factor analysis is that the data are multivariate normally distributed (Kline, 2011b). Prior studies with the MIC database showed that the univariate and consequently the multivariate distribution of many of the items in the MIC database are non-normal (Richardson et al., 2012). Because of the non-normality of the data, the models in stages one and two were estimated using a Robust Maximum Likelihood (MLR) estimator, given that the MLR estimator is reasonably robust against non-normality (Li, 2016a, b). For model estimation, a biased sample covariance matrix (a covariance matrix without Bessel's correction) was used as input (Rosseel, 2014).

Model fit was examined with the scaled versions of the χ2, Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Root Mean Squared Error of Approximation (RMSEA) and Standardized Root Mean Residual (SRMR) fit indices for models estimated with an MLR estimator. For Diagonally Weighted Least Squares (DWLS) estimated models, the non-scaled χ2, CFI, TLI, and the RMSEA fit indices were used to assess model fit. For the CFI and the TLI, a value higher than 0.900 was used as an indication of acceptable fit (Bentler and Bonett, 1980), and a value of 0.950 was used as a reference point for good fit (Hu and Bentler, 1999). The reference point of good fit for RMSEA was a value lower than 0.6, with a value lower than 0.8 indicating reasonable fit (Browne and Cudeck, 1992; Hu and Bentler, 1999). The reference point of good fit for the SRMR was a value lower than 0.8 (Hu and Bentler, 1999). In the base analysis, missing data were managed with a Full Information Maximum Likelihood (FIML) estimator for the models estimated with an MLR estimator. Missing data of the models estimated with the DWLS estimator were handled through pairwise deletion. The scales of the latent variables were fixed by fixing the first factor loading of a latent variable to one.

The robustness of the model fit estimates of the final measurement model itself was tested in three stages. In the first stage, this model was validated with the test dataset. In the second stage, further robustness tests were conducted by treating items with up to seven response options as categorical variables. The model was estimated using a (DWLS) estimator. Polychoric correlations were calculated for items that were treated as categorical variables. This model was estimated separately in the training and the test dataset. In the third stage, the effect of alternative ways of handling missing data on model fit was tested. The effect of a list-wise deletion and a pair-wise deletion of cases on model fit was tested with the model estimated with an MLR estimator. The

effect of a list-wise deletion of cases was tested with the model estimated with a DWLS estimator.

4.2.4 Instrument development

The full measurement model was reduced to a parsimonious instrument that was still sensitive to the different types of constructs that were included in the full measurement model. The instrument was developed in four steps.

1. The inter-factor correlations of the full measurement model were studied. If inter-factor correlations were higher than 0.9, then one factor was deleted. In this case, it was assumed that either factor was sufficiently sensitive to the information of the other factor (Le et al., 2010).
2. Items that cross-loaded on multiple factors were removed to ensure that the scale scores reflect single constructs.
3. Item-rest correlations and the effect of dropping an item on the reliability of an instrument were studied. Items that showed a particularly low item-rest correlation compared to other items' item-rest correlations were subject to further scrutiny. Also, items were further scrutinized if dropping them resulted in a material increase in reliability. Hard cut-off values were not used in these analyses, rather the statistical properties of single items were compared with each other to identify potential candidates for removal. Whether the items that were further scrutinized were included in the instrument depended on their content; if items existed that covered similar content, the decision was made to drop the relatively ill-functioning item.
4. Single items from a group of items that measured a similar construct were retained for the scale. For instance, when four items measured general satisfaction with life, only one item was selected for the instrument. The selection of these items was conducted on a case-by-case basis, but the general criteria that guided item selection were their item difficulty properties. To clarify, item difficulty was studied by observing the response patterns of items, with less skewed items being preferred over heavily skewed items with strong floor or ceiling effects. Furthermore, items were selected to minimize floor and ceiling effects. Lastly, in some cases the items' content was taken into account in the decision.

A detailed explanation of the item selection process can be found in the Appendix in Section 0. After reducing the number of items according to the steps presented above, a measurement model was estimated with a MLR estimator to test the model fit of the remaining items, which together form an instrument. The model was estimated with the full MIC database. Fit was assessed according to the fit indices introduced above (scaled versions of the χ2, CFI, TLI, SRMR, and the RMSEA fit indices).

One last thing I would like to note before presenting the results is that from this point onward individual items will be discussed. These items will be named according to their codes in the MIC study database (for example sf1), given that individually writing out each item would lead to a problematic read of this chapter. In Appendix Table 4 the item codes are linked to the full written-out version of the items, with an example of their respective response options.

4.3 Results

4.3.1 Data

A total of 9665 subjects participated in the MIC study (Richardson et al., 2012). After editing the data to remove unreliable responses, the database consisted of 8022 observations. For the creation of a training and a test dataset, this sample was split into two groups of 4011 observations. Further demographic information about the dataset can be found is presented in Table 4.

Of the 8022 observations in the complete dataset, 1191 contained missing values. These missing values mostly came from the Norwegian sample (n = 1177), because the ICECAP-A and the AQoL-4D instruments were not administered in this country. Given that these instruments were not administered to the participants in Norway, I handled this data as missing at random, given that the outcome variable (the assessment wellbeing) is not a cause that this data is missing.

Table 4. Demographics of the MIC study sample

Descriptive variables	***N* (%)**
Age group subsample	
18-24	513 (6.39%)
25-34	944 (11.77%)
35-44	1137 (14.17%)
45-54	1689 (21.05%)
55-64	2008 (25.03%)
65+	1731 (21.58%)
Gender subsample	
Male	3848 (47.97%)
Female	4174 (52.03%)
Education subsample	
High school	2522 (31.44%)
Some post-secondary, post-secondary certificate or diploma	3241 (40.40%)
University degree and higher	2259 (28.16%)
Health condition subsample	
Healthy public	1760 (21.94%)
Arthritis	929 (11.58%)
Asthma	856 (10.67%)
Cancer	772 (9.62%)
Depression	917 (11.43%)
Diabetes	924 (11.52%)
Hearing problems	832 (10.37%)
Heart problems	943 (11.76%)
Stroke	23 (0.29%)
Chronic obstructive pulmonary disease	66 (0.82%)
Country	
Australia	1430 (17.83%)
Canada	1330 (16.58%)
Germany	1269 (15.82%)
Norway	1177 (14.67%)
United Kingdom	1356 (16.90%)
United States of America	1460 (18.20%)
Total	8022 (100%)

Table published previously in Ubels et al. (2022c).

4.3.2 Sub-aim (1): Selection of items

The result of the item selection process can be found in Appendix Table 4. In this table, an overview of the themes, subthemes, their associated quotes, and the items linked to those quotes are provided. A total of 56 items were selected. Of these 56, 26 items were linked to the theme “Option Wellbeing”. A set of 20 items were linked to the theme “Perceived Access to Options”. An additional 6 items were linked to the theme “Self-Realization”. A last set of 4 items were linked to the theme “Perceived Control”. For some of the subthemes, no items could be identified. These subthemes were “Access due to Social Wellbeing”, ”Access due to Activity Wellbeing”, and “Access due to

Finances" from the theme "Perceived Access to Options". Also, no items could be linked to the subtheme "Having Dignity" from the theme "Self-Realization".

4.3.3 Sub-aim (2): Model development

Table 5 presents the fit index values of the various models that were tested in the confirmatory factor analysis. The analysis with the measurement model that was based on the theoretical framework developed in the last chapter indicated inadequate model fit (model 1 in Table 5). Theory-based respecifications were made to the model to study if model fit could be improved (an explanation about these respecifications can be found in Appendix Section 10.3). The respecifications resulted in improved, but still inadequate, model fit (model 2 in Table 5). Therefore, I decided to conduct a data-driven explorative analysis to study whether model fit could be improved.

Table 5. Fit statistics of tested models

	Dataset	χ^2	df	RMSEA ** (90% CI)	SRMR ***	CFI†	TLI†
Model 1: A priori model	training dataset	42,967.3	1,478	0.096 (0.095 – 0.097)	0.096	0.731	0.719
Model 2: After theoretical re-specifications	training dataset	30,079.2	1,462	0.080 (0.079 – 0.081)	0.090	0.816	0.806
Model 3: After post-hoc adjustments with method factors for response option length	training dataset	9,661.9	1,368	0.044 (0.043 – 0.045)	0.044	0.946	0.942
Model 4: Final model without method factors for response option length	training dataset	10,784.0	1,379	0.047 (0.046 – 0.048)	0.044	0.939	0.934
Model 5: Final model MLR*	test dataset	10,798.9	1,379	0.047 (0.046 - 0.047)	0.045	0.939	0.935
Model 6: Final model DWLS*	test dataset	16,232.2	1,379	0.052 (0.051 – 0.053)	0.045	0.964	0.961

Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Root Mean Squared Error of Approximation (RMSEA), Standardized Root Mean Residual (SRMR), degrees of freedom (df)
* Models 1 to 5 were estimated with a Robust Maximum Likelihood estimator (MLR), Model 6 with a Diagonally Weighted Least Squares (DWLS) estimator.
** Values lower than 0.6 indicate of acceptable fit.
*** Values lower than 0.8 indicate of acceptable fit.
† Values higher than 0.9 indicate acceptable fit.
Table published previously in Ubels et al. (2022c).

Based on this explorative analysis, the constructs related to subjective wellbeing were restructured. In the development of the a priori measurement model, items from the MIC database related to life satisfaction as well as certain emotional experiences were linked to the construct “Option Wellbeing”. However, the data-driven explorative analysis showed that model fit improved when items related to these emotional experiences were modeled as an independent factor. Further improvements in model fit were achieved by loading the satisfaction related items on another factor. Due to the changed nature of these factors, I decided to rename them into “Affective Wellbeing”, which represents emotional elements related to experienced wellbeing, and “Reflective Wellbeing”, which represents cognitive elements related to experienced wellbeing.

Furthermore, seven orthogonal-specific factors were created that were based on an inspection of the modification indices and residual correlations between the items. Factor (1) accounted for covariance amongst items that inquire about social aspects. Factor (2) accounted for covariance amongst items that covered the participants’ need for support or help. Factor (3) accounted for covariance amongst items that measure elements of happiness. Factor (4) accounted for covariance amongst items that covered the experience of anxiety. Factor (5) accounted for the covariance that was a result of the testlet layout of the items inquiring about physical limitations. Factor (6) accounted for the covariance that resulted from the testlet layout of the items that inquired about emotional limitations. I decided to let factor (5) and factor (6) correlate with each other, due to the similarities in the wording and presentation of these testlets, which makes these factors not strictly orthogonal. Factor (7) accounted for similarities amongst the items of the SF-36 V2 that inquired about negative emotional experiences, such as experiences of depression or feeling down. In the SF-36 V2, these items are part of the “Emotional Wellbeing” subscale.

Besides the construction of orthogonal factors, the error terms of some items were also correlated to account for unexplained variance. The error terms of items PWI_c and sf1 were correlated with each other since both ask participants to provide an overall rating of their own health. Also, the error terms of the items aqol24 and sf22 were correlated with each other, since both ask whether pain influences individuals’ normal activities or work. The error terms of the items sf20 and sf32 were correlated with each other, since both inquire whether health or emotional problems hinder social activities.

Lastly, the items aqol2_4D and aqol30 were correlated with each other, since both cover whether individuals need help with their household tasks.

Another category of adjustment concerned the relation between factors and items. Some items were respecified to cross-load on multiple factors. Items sf17, sf18, and sf19 cover whether individuals consider their emotional state to have limited their ability to work or conduct activities. Besides loading on the “Perceived Access to Options” factor, these items were allowed to cross-load on the “Affective Wellbeing“ factor. The loadings of other items were respecified altogether. The item aqol5_4D covers the experience of loneliness. The explorative data analysis showed that model fit was improved if this item loaded on the factor “Affective Wellbeing” instead of the factor “Reflective Wellbeing”. The measurement model was changed accordingly. The item aqol10, which covers how satisfying and pleasant close relationships are, was also respecified to load on the factor “Affective Wellbeing”. The most notable change however concerned three items from the ICECAP-A instrument, which is developed to measure capabilities: ic03 (ability to be independent), ic04 (ability to achieve and progress), and ic05 (ability to be happy). Initially, the items ic03 and ic04 loaded on the “Self-Realization” factor. Item ic05 loaded on the “Option Wellbeing” factor. The residual correlation matrix however showed that model fit could be improved if these three items loaded on the “Perceived Control” factor. Furthermore, item ic03 was respecified to load on the “Perceived Access to Options” factor. Item ic05 was respecified to load on the factor “Affective Wellbeing”.

One item was deleted from the model. This item, ONSk, measures the “happiness” of individuals. However, the item correlated strongly with the items from the “Reflective Wellbeing” factor instead of the “Affective Wellbeing” factor. I hypothesized that this unexpected correlation could be explained by the layout of the MIC survey, where the item ONSk follows the items ONSi and ONSj. ONSi and ONSj both cover elements of “Reflective Wellbeing”, respectively measuring life satisfaction and the experience of doing something worthwhile. This could have “anchored” responses to the item ONSk. Therefore, I decided to delete item ONSk, given that the item did not reflect the construct that it is supposed to measure.

During the explorative analysis, I also developed a model with method factors that accounted for covariance between items related to the length of response options of

items. Initially, a solution for this model could not be estimated. Still, potentially other method factors could exist related to the layout of the MIC survey and similarities between items within specific questionnaires. Particularly the ICECAP-A, the SWLS, and the PWI instruments were susceptible to these sources of additional correlation. In the case of the ICECAP-A, the three items included in the model have similar wording, which covers what an individual “can” or “is able” to do. The items in the model from the SWLS were all scored on a seven-point scale, from strongly disagree to strongly agree. Similarly, the items of the PWI which were included in the model were all scored on an eleven-point scale, from completely dissatisfied to completely satisfied. Therefore, three orthogonal factors were created which aimed to explain the covariance that is a result of the ways that items were presented: one for the items of the PWI, one for the items of the SWLS, and one for the items of the ICECAP-A. The fit indices of this model are presented as Model 3 in Table 5.

Model 3 was however complex and the items showed relatively low loadings on their respective method factors. For the PWI factor, standardized loadings ranged from 0.155 to 0.317, with 4 out of 5 items having a standardized factor loading lower than 0.300. For the SWLS factor, standardized factor loadings ranged from 0.228 to 0.286. For the ICECAP-A factor, standardized factor loadings ranged from 0.186 to 0.602, with 2 out of 3 items having a standardized factor loading below 0.300. Therefore, a model was developed without method factors. This model is essentially a nested model of Model 3. Fit indices of this model can be found under Model 4 in Table 5. A comparison of the fit indices between models 3 and 4 shows comparable fit, with differences in CFI, SRMR, and RMSEA being smaller than 0.01, 0.015, and 0.015 respectively. According to Chen (2007), such small differences in nested models show that the added constraints do not provide a significant improvement to the model. Since the measurement model of Model 4 is more parsimonious it was decided to retain it as the final measurement model that was retained for further analysis with the test dataset. This analysis resulted in Model 5. Table 6 shows the factor loadings, Table 7 presents the item-error correlations of paired items, and Table 8 the inter-factor correlations of Model 5.

Table 6. Standardized factor loadings per factor of Model 5

	*Reflective wellbeing**	*Affective wellbeing**	*Perceived access to options**	*Perceived control**	Social aspects	Need for help or support	Anxiety	Happiness	SF-36 negative emotions	SF-36 phys testlet†	SF-36 emo testlet‡
PWI_a	0.863										
PWI_c	0.458		0.414								
PWI_d	0.829										
PWI_e	0.639				0.339						
PWI_g	0.613										
ONSi	0.928										
ONSj	0.813										
SWLS_a	0.870										
SWLS_c	0.911										
SWLS_d	0.784										
aqol27	0.405	0.526						0.157			
sf1	0.285		0.555								
ONSl		0.448					0.410				
sf17		0.432	0.303								0.684
sf18		0.505	0.241								0.644
sf19		0.437	0.285								0.653
sf20		0.440	0.418								
sf24		0.645					0.368		0.339		
sf25		0.750							0.444		
sf28		0.822							0.321		
sf29		0.484	0.257						0.222		
sf30		0.724						0.302			
sf32		0.399	0.465								
aqol5		0.847									
aqol9		0.247	0.586								
aqol10		0.633			0.578						
aqol18		0.782									
aqol20		0.807						0.353			

	*Reflective wellbeing**	*Affective wellbeing**	*Perceived access to options**	*Perceived control**	Social aspects	Need for help or support	Anxiety	Happiness	SF-36 negative emotions	SF-36 phys testlet†	SF-36 emo testlet‡
aqol23		0.541			0.602						
aqol25		0.689						0.368			
aqol33		0.896									
aqol34		0.576			0.365						
aqol35		0.816									
aqol5_4D		0.678			0.238						
aqol11_4D		0.839					0.201				
aqol1_4D			0.613			0.720					
aqol2_4D			0.729								
aqol3_4D			0.665			0.309					
aqol3			0.842								
aqol4			0.755								
aqol15			0.799								
aqol19			0.782			0.203					
aqol24			0.761								
aqol26			0.301	0.515							
aqol30			0.829								
sf13			0.700							0.540	
sf14			0.716							0.549	
sf15			0.762							0.558	
sf16			0.768							0.554	
sf22			0.773								
ic03			0.379	0.314							
aqol21				0.817							
aqol29				0.823							
ic04				0.741							
ic05				0.622							

*Columns in italic are the main constructs of interest. Other factors represent specific constructs that are added to account for covariance between items. The items linked to the codes can be found in Appendix Table 4. † SF-36 Phys Testlet: SF-36 Physical Limitations Testlet; ‡ SF-36 Emo Testlet: SF-36 Emotional Limitations Testlet. Table published previously in Ubels et al. (2022c).

Table 8 shows that the factors generally show a positive correlation. Two factors are highly correlated: the correlation between “Perceived Control” and “Affective Wellbeing” is 0.911. Other factors showed correlations ranging from 0.395 to 0.756.

Table 7. Item error correlations of Model 5.

Items	**Error correlation value**
PWI_c – sf1	0.489
aqol24 – sf22	0.458
sf20 – sf32	0.425
aqol2_4D – aqol30	0.304

The items linked to the codes can be found in Appendix Table 4.
Table published previously in Ubels et al. (2022c).

Table 8. Standardized inter-factor correlations of Model 5

Factors	Reflective wellbeing	Affective wellbeing	Perceived control	Perceived access to options	Physical limitations testlet	Emotional limitations testlet
Reflective wellbeing	1					
Affective wellbeing	0.750	1				
Perceived control	0.764	0.918	1			
Perceived access to options	0.380	0.512	0.594	1		
Physical limitations testlet	0	0	0	0	1	0
Emotional limitations testlet	0	0	0	0	0.468	1

Note: factors that are not part of this table are orthogonal.
Table published previously in Ubels et al. (2022c).

The various robustness tests showed that the model fit results were generally robust. Estimating the model with the DWLS estimator resulted in an improved fit in terms of the CFI, TLI, and SRMR fit indices, with a slight reduction in fit in terms of RMSEA when missing data were handled with a list-wise deletion of cases (see Table 5). Alternative ways of handling missing data did not result in meaningful changes in model fit.

4.3.4 Sub-aim (3) Instrument development

From the 56 items from the full measurement model, an instrument was developed that consists of 15 items. A detailed explanation of this selection process can be found in

the Appendix in Section 10.3. Since this instrument aims to cover wellbeing and is conceptually based on “option freedom”, I decided to give it the name the Wellbeing Related option Freedom (WeRFree) instrument. This instrument can be found in the Appendix in Section 10.3.6.

The WeRFree instrument thus consists of three scales to measure wellbeing. Experienced wellbeing is measured with the “Reflective Wellbeing” scale (Cronbach’s alpha: 0.89, 6 items) and the “Affective Wellbeing” scale (Cronbach’s alpha: 0.83, 4 items). Health-related capabilities are measured with the “Perceived Access to Options” (Cronbach’s alpha: 0.88, 5 items). Figure 3 shows a graphic representation of the measurement and structural model of the WeRFree instrument. The 15 items of the WeRFree instrument were specifically selected to capture the various constructs with parsimonious scales that have minimal floor – and ceiling effects. The ceiling effects of the “Perceived Access to Options”, “Reflective Wellbeing”, and “Affective Wellbeing” scales were 16.85%, 0.96%, and 1.62% respectively. The floor effects of the “Perceived Access to Options”, “Reflective Wellbeing”, and “Affective Wellbeing” scales were respectively 0.05%, 0.04%, and 0.17%. Table 9 gives an overview of the standardized factor loadings, standardized intercepts, and standardized variances of the items of the WeRFree instrument on their respective scales. The fit indices indicated good model fit ($\chi 2 = 1756.82$, df = 105, CFI = 0.970, TLI = 0.963, RMSEA = 0.055, SRMR = 0.036). Appendix Table 5 gives an overview of the proportion of responses per response option per item of the WeRFree instrument. Appendix table 6 presents item-rest correlations per scale.

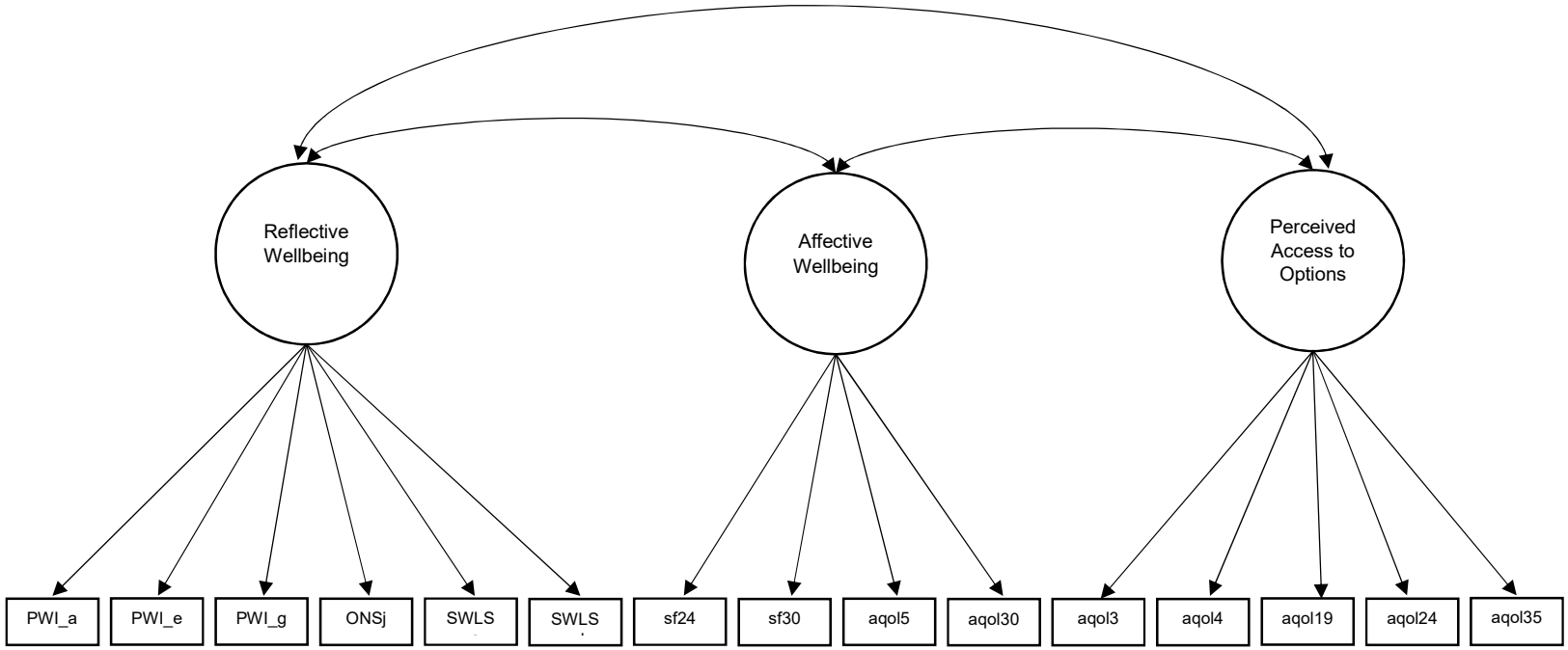


Figure 3.Graphical presentation of the WeRFree measurement model

Note: The item linked to the codes can be found in Appendix Table 4. The full instrument is also presented in the Appendix in Section 10.3.6. Figure has been published in Ubels et al. (2022c).

Table 9. Standardized statistics of the WeRFree instrument

Construct **Item**	**Standardized loadings**	**Standardized intercepts**	**Standardized variances**
Reflective Wellbeing			
PWI_a	0.851	3.155	0.276
PWI_e	0.693	2.928	0.520
PWI_g	0.661	2.964	0.563
ONSj	0.824	3.011	0.322
SWLS_a	0.839	2.602	0.295
SWLS_d	0.785	2.915	0.385
Affective Wellbeing			
sf24	0.627	3.926	0.607
sf30	0.725	3.335	0.459
aqol5	0.830	3.846	0.312
aqol35	0.813	4.193	0.340
Perceived Access to Options			
aqol3	0.848	4.265	0.281
aqol4	0.754	4.177	0.431
aqol19	0.788	6.015	0.380
aqol24	0.743	3.503	0.448
aqol30	0.812	4.552	0.341

The items linked to the codes can be found in Appendix Table 4. The full instrument is also presented in the Appendix in Section 10.3.6.
Table published previously in Ubels et al. (2022c).

Table 10. Inter-factor correlations of the WeRFree instrument

Factors	Reflective wellbeing	Affective wellbeing	Perceived access to options
Reflective wellbeing	1	-	-
Affective wellbeing	0.762	1	-
Perceived access to options	0.406	0.560	1

Table published previously in Ubels et al. (2022c).

4.4 Discussion

The study presented in this chapter had three aims. The first aim was to identify items in the MIC study database that could be used to measure the four themes (“Perceived Access to Options”, “Perceived Control”, “Option Wellbeing”, and “Self-Realization”) that emerged from the analysis in Chapter 3. The second aim was to conduct a confirmatory factor analysis to establish whether the themes can be operationalized as constructs. The third aim was to develop an instrument.

4.4.1 Sub-aim (1): Selection of items

The first aim was to identify relevant items from the MIC database for the measurement of the themes and subthemes. This aim has been fulfilled with partial success. The MIC database contained items that were relevant for most of the subthemes. However, relevant items could not be identified for the subthemes Access due to Social Wellbeing", "Access due to Activity Wellbeing", "Access due to Finances", and "Having Dignity". Further research is necessary to examine whether items related to these elements should be added to the WeRFree instrument. This can be done by studying the content validity of the instrument with potential users, such as patients. This will provide information about whether the instrument comprehensively captures all the relevant elements of capability in the context of HTA.

4.4.2 Sub-aim (2): Model development

The second aim was to study whether the themes from the qualitative framework developed in Chapter 3 could be measured as individual constructs. To do this, an a priori measurement model was developed. This model used items from the MIC database. The high degree of model misfit indicated that the a priori specified measurement model did not predict the data well, which meant that the internal structure of constructs was different than hypothesized. This led to a data-driven explorative restructuring of the measurement model, to study where the theoretical model can be improved.

As a result of this explorative analysis, the subjective experience of capability is measured with the constructs "Affective Wellbeing" and "Reflective Wellbeing". this structure of subjective wellbeing closely follows insights from psychology, in which subjective wellbeing is conceptualized as consisting of cognitive judgements and of emotional experiences (Diener, 1984). In psychology, the affect construct is usually subdivided in separate factors representing positive and negative emotions (Busseri and Sadava, 2011; Diener, 1984; Lucas et al., 1996).

A further notable diversion between the exploratory data-driven developed model and the a priori model concerns three items from the ICECAP-A instrument (Al-Janabi et al., 2012): the items ic03 (covers the ability to be independent), ic04 (covers the ability to achieve and progress) ic05 (covers the ability to be happy). These items are

developed to measure various kinds of capabilities that people consider to be important in their lives (Al-Janabi et al., 2012). Interestingly, each of these three items (cross-) loaded on the “Perceived Control” factor. These items were developed to assess capability as objectively as possible, with the wording “I am able to…” or “I can…” Similar wording has been used in the field of psychology to measure concepts such as “self-efficacy” (Frei et al., 2009). Self-efficacy is a construct that covers the individual’s perception of being able to have control over their own functioning and their environment (Bandura, 2001). In psychology, self-efficacy is seen as an important determinant of subjective wellbeing, but is not part of the subjective wellbeing construct itself (Bandura, 2008). Other research has identified the similarities between instruments inquiring about self-efficacy and the ICECAP-A as well (Rohrbach et al., 2021). An interesting question for future research might be whether existing perceived self-efficacy instruments could be used as an indicator for the measurement of capabilities.

The allocated items to the construct “Perceived Access to Options” in the a priori model was largely similar to the data-driven developed model. This construct represents the perceived health-related capabilities of individuals. In the measurement model, this construct was measured with items that cover physical and mental health-related difficulties that affect the ability of individuals to access a range of options, such as mobility, self-care, or the individual’s ability to conduct their social responsibilities. These items were derived from HRQoL instruments that are used in contemporary health economics. Consequently, these instruments seem to some extent be able to assess the health-related capabilities of individuals. This has also been argued by Cookson (2005).

Besides the measurement model, further observations can be made regarding the structural model. The factors correlations of the factors that were allowed to correlate with each other were generally lower than 0.9 (between “Reflective Wellbeing” and “Perceived Control”, “Reflective Wellbeing” and “Affective Wellbeing”, as well as “Perceived Access to Options” and the three other factors). This indicates that from a measurement perspective, these constructs can be differentiated from each other and each construct provides unique information that can be used for research or decision-making (Kline, 2011c). Two factors showed a correlation that was higher than 0.9: “Perceived Control” and “Affective Wellbeing”, which indicates that the constructs are

largely overlapping from a measurement perspective (Le et al., 2010). This is not a surprising result in light of psychological literature. Studies show that there is a (strong) relationship between self-efficacy and subjective wellbeing (Azizli et al., 2015; Lent et al., 2005; Strobel et al., 2011). Furthermore, studies have shown that an individual's level of self-efficacy mediates the effect of physical and mental health problems on subjective wellbeing (De Castro et al., 2012; Marks and Allegrante, 2005; Martinez-Calderon et al., 2020; Schönfeld et al., 2016). The correlations that were identified in the structural model are thus supported by findings of other studies.

In this chapter, a confirmatory factor analysis was conducted to conduct tests and identify improvements. An alternative method that could have been used is exploratory factor analysis. In an exploratory factor analysis, the factors are identified through the use of statistical heuristics (Costello and Osborne, 2005). The main reason for not applying exploratory factor analysis in this study is that I suspected that certain items shared variance (e.g., multiple items sharing similar content, the presence of testlets) that was not of interest for the study's main study aim: the operationalization of the themes from Chapter 3 into constructs. As a consequence, the factors that would be the result of such an exploratory factor analysis would have been not useful for this study.

Therefore, I conducted a confirmatory factor analysis that, amongst other things, resulted in the addition of specific factors to explain covariance between certain sets of items. As mentioned, the use of orthogonal-specific factors is a key element of bifactor modeling. Bifactor models have been used in a variety of different settings, one of them is to control for multi-dimensionality (Rodriguez et al., 2016). In this context, the function of the specific factors is to control for item correlations that are not of direct interest to the researcher (Rodriguez et al., 2016). Similarly in this chapter, item correlations that were unaccounted for by the constructs of interest were modeled by specific factors.

4.4.3 Sub-aim (3): Instrument development

The third aim of this study was to develop an instrument that can be used to assess wellbeing. As was mentioned in the results, I decided to not include a scale to measure the construct "Perceived Control" in the WeRFree instrument. The choice to include a scale to measure "Affective Wellbeing" instead of a scale to measure "Perceived

Control" was based on theory. In Chapter 3, I explained that the "Perceived Control" construct influences the relationship between "Perceived Access to Options" and the experienced wellbeing constructs, which are represented in the WeRFree instrument as "Reflective Wellbeing" and "Affective Wellbeing". Consequently, the "Perceived Control" construct is not an outcome variable per se, since it mediates or moderates the relationship between the "Perceived Access to Options" and experienced wellbeing constructs. Given that the "Affective Wellbeing" construct was an outcome in this constellation, I decided to not include a scale to measure "Perceived Control".

As this chapter shows, developing a capability instrument that is based on the concept of option freedom has two advantages over existing capability instruments that have been based on Sen's concept of capability. First, the WeRFree instrument shows how the concept of option freedom not only facilitates the identification of various capabilities that are relevant to individuals' wellbeing, but also in identifying the (health-related) difficulties that individuals might experience while trying to achieve those capabilities. Second, the clarity of the option freedom concept facilitates the differentiation between items that reflect capability and items that reflect other elements that are key to wellbeing, such as functionings. The latter advantage might improve the assessment of wellbeing, since theoretically certain elements of wellbeing are more appropriately assessed in terms of functionings than capabilities (such as subjective wellbeing, see Section 2.4 of this dissertation).

4.4.4 Limitations

One limitation that was already mentioned in this discussion is that items could not be identified for all the subthemes of the theoretical framework developed in Chapter 3. An additional imitation is the explorative, data-driven nature of this study. This affects the interpretation of the fit index values since they are statistical tests that are developed for hypothesis testing. Data-driven adjustments based on these fit indices result in overly optimistic estimates of model fit. (Wagenmakers et al., 2012). I tried to minimize this limitation by dividing the dataset from the MIC study randomly into a training dataset and a test dataset. Explorative analysis and data-driven adjustments of the measurement model were conducted with the training dataset and the final version of the measurement model was tested on the test dataset. Still, validating a model with this approach is not ideal, since measurement errors that are caused by research design that might have affected the data-driven choices to improve the model

are not controlled for. Consequently, model fit could still be overestimated. Further validation of the model and the WeRFree instrument should therefore be based on external datasets.

4.4.5 Conclusion

To conclude, in this chapter I have developed an instrument based on the concept of option freedom (Pettit, 2003): the WeRFree instrument. The WeRFree instrument needs to be understood as a proposal of how this concept can be operationalized in the context of measuring health-related capabilities. Furthermore, the instrument provides evidence for measuring wellbeing in terms of both functionings related to subjective wellbeing and capabilities. Still, the WeRFree instrument will require further development in before it can be used in practice to assess the wellbeing of individuals, for example in the context of HTA.

5 MEASUREMENT INVARIANCE AND ADAPTED PREFERENCES

5.1 Background

As discussed in Chapter 1, adapted preferences form a challenge for the assessment of wellbeing with self-report instruments (see Section 1.5.3 for a discussion about adapted and adaptive preferences). Indeed, several authors have questioned whether people are informed enough to judge their own health-related wellbeing because of this phenomenon (Culyer, 1989; Hurley, 2000; Sen, 2002). In the context of the WeRFree instrument, this means that responses to items could potentially be affected by the adapted preferences of individuals (see for examples in HRQoL instruments Huang et al., (2011), Knott et al., (2017b), Smith et al., (2016)). Consequently, the same item might have different meanings for different individuals, which provides a challenge for the interpretation of results of self-reported wellbeing instruments.

Besides forming a challenge for wellbeing assessment of individuals, such differences in the interpretation of items also affect comparisons of instrument scores between groups. For HTA, the latter problem is particularly important, due to the reliance of health economic analysis on self-report instruments to assess their effects (Coast et al., 2018; Groot, 2000; Knott et al., 2017a). Adapted preferences might lead to an over – or underestimation of the effects of these technologies, which results in an inadequate value assessment (Groot, 2000; Knott et al., 2017a; Mitchell et al., 2015).

As an example of how adapted preferences might affect HTA, consider a new health technology that improves mobility. In this case, it might be difficult to measure its real effect when individuals with limited mobility already report having a high initial level of mobility before the use of such a health technology (Knott et al., 2017a). This could lead to an unjust allocation of resources if, due to adapted preferences, the information that policymakers receive indicates that a new health technology only has a minor effect (Groot, 2000; Knott et al., 2017a; Mitchell et al., 2015). In the context of measuring the effectiveness of health technologies, this means that the effect of a new health technology in a patient group with adapted preferences might be underestimated because of adapted preferences (Knott et al., 2017a).

Thus, adapted preferences can affect how individuals interpret and respond to items. It is therefore important to test whether different groups interpret items similarly. One

way of doing so is by testing for measurement invariance. Measurement invariance has been defined by Millsap (2007) as follows: “some properties of a measure should be independent of the characteristics of the person being measured, apart from those characteristics that are the intended focus of the measure”.

In the context of health economics, this means that instruments measure the constructs that they intend to measure, such as capability wellbeing or HRQoL, without interference from other characteristics, such as age, level of education, or adapted preferences. In other words, the instrument scores should only reflect changes in the constructs of interest. With respect to adapted preferences, it should therefore be possible to conduct a measurement invariance test to study whether advantaged and disadvantaged populations interpret items similarly. The establishment of measurement invariance would then indicate that item responses are not affected by adapted preferences.

The measurement invariance properties of instruments have been tested in the context of cross-cultural research (Jang et al., 2017; Jeong and Lee, 2019). In the context of quality-of-life instruments, measurement invariance testing has also been used to study whether responses to instruments change over time in patient groups (Sajobi et al., 2018). Such tests have however not been systematically performed in applications of the capability approach in health economics. Recent reviews of the psychometric properties of capability instruments did not identify this type of evidence for these instruments (Helter et al., 2020; Till et al., 2021). Only one recent publication by Rencz et al. (2021) was identified in which the measurement invariance properties of a capability instrument have been studied. Amongst other things, this study tested the measurement invariance properties of the ICECAP-A in different subgroups in a sample of dermatological patients (Rencz et al., 2021). Measurement invariance could not be established in subgroups where participants were categorized according to age, marital status, or scores on a dermatology-specific QoL index. However, adapted preferences were not considered as a possible cause for the differences in interpretation of items in these groups.

I also identified one further qualitative study that aimed to assess whether responses to the ICECAP-A, ICECAP-SCM, and the EQ-5D-5L were influenced by adapted and adaptive preferences by means of think-aloud interviews (Coast et al., 2018). The

conclusion of this study was that there is little indication of adaptive preferences in an end-of-life setting (Coast et al., 2018). Although this study provides an important qualitative insight into this particular group's reasoning when responding to items, it is unclear if these responses are quantitatively comparable across groups.

Therefore, the primary aim of this chapter is to establish whether the WeRFree instrument is measurement invariant across groups of individuals that differ in terms of age, education, gender, or health condition. Once measurement invariance was established, the secondary aim was to descriptively compare the differences between scale scores between these different groups.

5.2 Methods

5.2.1 Data

For the study in this chapter, the MIC study database was used. The general characteristics of the MIC study database have been described in the last chapter. With respect to the study presented in this chapter, the complete sample was grouped into different subsamples. These subsamples were grouped according to age, level of education, gender, and health condition (see Table 4 in Chapter 4).

5.2.2 Using the WeRFree instrument for measurement invariance testing

The WeRFree instrument was used for the study presented in this chapter. Item scores of the WeRFree instrument were rescaled to produce scale scores that range from 0 - 100, with a score of 100 meaning the optimal level of capability or subjective wellbeing in each of the scales. To ensure that each of the items has a similar weight on their respective scale, I adjusted item scores for response option length. For example, the score of a respondent on an item with 11 response options was divided by 11 and then multiplied by 100. The means and standard deviations for these scores were calculated.

5.2.3 Statistical analyses

5.2.3.1 Measurement invariance

A multi-group confirmatory factor analysis was conducted to test for four different types of measurement invariance: (1) configural invariance, (2) metric invariance, (3) scalar invariance, and (4) residual invariance (Chen, 2007; Meredith and Teresi, 2006;

Widaman and Reise, 1997). Essentially, for each of these analyses, a different model is produced that is progressively more restrained than the last model.

An instrument is configural invariant if the three-factor structure of the WeRFree instrument can be identified in different groups. To illustrate why this is important, one can imagine an instrument that can be used to measure emotions that consist of seven items. These seven items are distributed over two scales. Three items reflect positive emotions, three items negative emotions, and one item asks about whether an individual tears up frequently. It could be the case that some groups only tear up due to sadness, while other groups tear up because of laughter. When responding to this instrument, the factor structure between these groups might thus be different, since the “tear up” item has a different meaning in different groups and can therefore load on either of the two scales. Configural invariance can then not be established, since individual items are related to different constructs.

When configural invariance can be established, metric invariance can be tested (Chen, 2007; Widaman and Reise, 1997). An instrument is metric invariant when the factor loadings are invariant across different groups. As explained in Section 1.6.1 in the introduction of this dissertation, the factor loading represents the strength of the relationship between a construct and an item, or, in other words, in how far the response to an item increases per unit increase in the latent variable. Invariant factor loadings indicate that the constructs influence changes in item scores in the same way in different groups. Again, the example presented under configural invariance can be used to illustrate the importance of metric invariance. Imagine that in two groups the “tear up” item loads on the happiness scale. However, also imagine that these two groups represent samples from different countries: in one country it is normal to physically show emotions, while in another country one is expected to moderate one's expression. When administering the instrument, the item might show a high factor loading in the former group, but a low factor loading in the latter group. In fact, the factor loading might be so low in the latter group that it is hard to identify a relationship between the positive emotion construct and the item. This is again a sign that an item is interpreted differently in these countries, which also means that scale scores cannot be compared directly.

The third type of invariance that will be tested is scalar invariance. An instrument is scalar invariant when the intercepts of each item are the same across different health conditions. Once scalar invariance is established, it is possible to compare mean scores of the scales between the health conditions, given that the scores can be assumed to represent real changes in the latent constructs (Chen, 2007; Widaman and Reise, 1997). Again, one can take the “tear up” item to illustrate the importance of scalar invariance. It is possible that this item loads with equal strength on the positive emotions scale in different groups and still varies in its expression in different groups. To illustrate, one can imagine that two groups experience a similar level of happiness. However, due to cultural differences, one group may be in tears from laughter, while the other group may not be expressing their happiness at all because their threshold for laughter is different. In this case, item scores would be incomparable.

Lastly, I studied whether residual invariance (strict invariance) can be established. Essentially, this means that the residuals of the items are similar across different groups. This indicates that the mean differences in scale scores that can be observed between groups are a result of differences in the latent construct and not caused by other factors (Chen, 2007; Widaman and Reise, 1997). This provides extra confidence that the difference in mean scores is indeed driven by differences in the latent construct of interest and not by other unmeasured constructs (Chen, 2007; Widaman and Reise, 1997). In the case of the “tear up” item presented above, it could be that other variables influence the response of items, which results in higher variability in the item. Such factors might for example be differences in the variability in eating spicy food in different groups, which could influence the frequency of individuals tearing up.

The models were estimated with a ML estimator. The reason for choosing a ML estimator over a DWLS estimator that is better suited for ordinal data (see Chapter 4) was that in order to properly use a DWLS estimator with ordinal data, sufficient observations have to be made in each response option of each item. Unfortunately, this was not the case with some of the items included in the WeRFree instrument. To illustrate, only six respondents of the healthy public subgroup in the health condition subsample responded to the lowest two response options of item aqol19. The reason for not using a MLR estimator was that the changes in model fit were difficult to compare due to the scaling adjustments associated with such estimators. Given the evidence that a ML estimator behaves reasonably well when modeling variables with

five or more response options, which matches the items included in the WeRFree instrument (Bandalos, 2014; Rhemtulla et al., 2012), I decided to use a ML estimator. Missing data were handled with FIML estimations (Enders and Bandalos, 2001).

Various fit indices were used to establish measurement invariance. To study configural fit, I used the same fit indices with similar values as indications for acceptable fit as in the last chapter (Bentler and Bonett, 1980; Hu and Bentler, 1999): $\chi 2$, CFI (value higher than 0.900 indicating acceptable fit), TLI (value higher than 0.900 indicating acceptable fit), RMSEA (value lower than 0.08 indicating acceptable fit) and SRMR (value lower than 0.08 indicating acceptable fit). To study the other forms of measurement invariance, the suggested fit index values by Chen (2007) for group sizes that are equal to or larger than 300 were followed, because the sample sizes of the groups in the different subsamples are larger than 300. For these measurement invariance tests, I used the ΔCFI, the ΔRMSEA, and the ΔSRMR fit indices. A score ≥ -0.010 in ΔCFI, ≥0.015 in ΔRMSEA, and a score of ≥0.030 in SRMR indicated noninvariance regarding metric invariance. Scores of ≥-0.010 in ΔCFI, ≥0.015 in ΔRMSEA, and ≥0.010 in SRMR were used as an indication of noninvariance regarding scalar and residual invariance. The chi-square difference test was not used to compare model fit, because of the large sample sizes of the subsamples, which would result in trivial differences in model fit being flagged as significant (Chen, 2007).

5.2.3.2 Testing mean differences

In the gender subsample, an independent samples t-test was used to study whether the difference between the means was significant. A one-way analysis of variance (ANOVA) was conducted to study whether a significant difference in scale scores existed between the groups in the other subsamples (Kim, 2017). When a significant difference was identified with the ANOVA, a Tukey-Kramer test was applied to study which pairs of groups showed statistically significant differences (Driscoll, 1996; Kramer, 1956). Before conducting the t-test, ANOVA, and the Tukey-Kramer test, the distribution, as well as the variances of the scales were studied by means of boxplots and residuals versus fitted graphs respectively. Observations with missing data were deleted for this part of the analysis.

5.3 Results

Table 11 presents the sample size per subsample, as well as the size of different groups within those subsamples. The only difference between Table 11 and Table 4 of Chapter 4 is that the total size of the health condition subsample is slightly lower compared to the size of the other subsamples. This is a consequence of the deletion of two "artifact" disease groups. During the recruitment phase of the MIC study project, the Australian arm also recruited patients affected by stroke and chronic obstructive pulmonary disease. These subgroups consisted of 23 and 66 participants respectively. The sample sizes of these groups were considered to be inadequate for further analysis and the groups were not included in the health condition subsample. In addition, 15 observations in the MIC database had missing data for items included in the WeRFree instrument. For the "Reflective Wellbeing" scale, 14 observations were missing for all included items. For the "Perceived Access to Options" scale, sf24 and sf30 had a missing value.

Table 11. Sample size per group per subsample for measurement invariance testing

Subsample	***N* (%)**
Age group subsample	
18-24	513 (6.39%)
25-34	944 (11.77%)
35-44	1137 (14.17%)
45-54	1689 (21.05%)
55-64	2008 (25.03%)
65+	1731 (21.58%)
Gender subsample	
Male	3848 (47.97%)
Female	4174 (52.03%)
Education subsample	
High school	2522 (31.44%)
Some post-secondary, post-secondary certificate or diploma	3241 (40.40%)
University degree and higher	2259 (28.16%)
Total in age group, gender, and education subsamples	8022 (100%)
Health condition subsample	
Healthy public	1760 (22.19%)
Arthritis	929 (11.71%)
Asthma	856 (10.79%)
Cancer	772 (9.73%)
Depression	917 (11.56%)
Diabetes	924 (11.65%)
Hearing problems	832 (10.49%)
Heart problems	943 (11.89%)
Total in health condition subsample	7933 (100%)

Table 12 presents the results of the measurement invariance test. Configural invariance was established in every subsample: the highest value for the upper level of the RMSEA 90% confidence interval is reached in the health condition and age group subsamples with a value of 0.060. The highest SRMR value is 0.041 in the health condition subsample. The lowest CFI value is 0.963 can be found in the health condition subsample as well. Metric invariance was also established in every subsample. The largest reduction in model fit in terms of CFI and SRMR could be identified in the health condition subsample, with a reduction in fit values of 0.0031 and 0.008 respectively. Scalar invariance was also established in every subsample. The largest reductions in RMSEA and SRMR, both reduced by 0.004, were identified in the age groups subsample. Furthermore, a 0.0098 reduction in CFI was identified in the health condition subsample.

Residual invariance was established in the education and gender subsamples. Residual invariance could not be established in the age group and health condition subsamples. For the health condition subsample, model fit in terms of CFI, RMSEA, and SRMR deteriorated by 0.0547, 0.021, and 0.029 respectively when comparing the residual invariance factor model to the scalar invariance factor model. For the age group subsample, CFI deteriorated by 0.0105, which also meant that residual invariance could not be established. To conclude, full measurement invariance was not established in the health condition and age group subsamples: in these subsamples the WeRFree instrument was measurement invariant up to scalar invariance. In the other subsamples full measurement invariance was established.

Table 12. Measurement invariance per subsample

Subsample	Model	X^2 (*df*)	CFI	RMSEA (90% CI)	SRMR	Δ X^2 (Δ *df*)	Δ CFI	Δ RMSEA	Δ SRMR
Health condition	Configural invariance	2990.14 (696)	0.961	0.058 (0.056 - 0.060)	0.041	-	-	-	-
	Metric invariance	3257.81 (780)	0.958	0.057 (0.055 - 0.059)	0.049	267.67 (84)	-0.0031	0.001	0.008
	Scalar invariance	3916.74 (864)	0.948	0.060 (0.058 - 0.062)	0.052	658.93 (84)	-0.0098	0.003	0.003
	Residual invariance	7233.82 (969)	0.893	0.081 (0.079 - 0.082)	0.081	3317.07 (105)	-0.0547	0.021	0.029
Age	Configural invariance	2853.83 (522)	0.966	0.058 (0.056 - 0.060)	0.039	-	-	-	-
	Metric invariance	3004.33 (582)	0.965	0.056 (0.054 - 0.058)	0.043	150.50 (60)	-0.0013	-0.002	0.004
	Scalar invariance	3669.88 (642)	0.956	0.059 (0.058 - 0.062)	0.047	665.54 (60)	-0.0089	0.004	0.004
	Residual invariance	4465.45 (717)	0.945	0.063 (0.061 - 0.064)	0.051	795.58 (75)	-0.0105	0.003	0.004
Gender	Configural invariance	2367.76 (202)	0.969	0.056 (0.054 - 0.058)	0.036	-	-	-	-
	Metric invariance	2399.61 (215)	0.969	0.054 (0.053 - 0.056)	0.037	31.84 (12)	-0.0002	-0.002	0.001
	Scalar invariance	2705.98 (228)	0.965	0.056 (0.054 - 0.058)	0.039	306.37 (12)	-0.0041	0.002	0.002
	Residual invariance	3028.31 (244)	0.960	0.057 (0.056 - 0.059)	0.041	322.33 (15)	-0.0043	0.001	0.002
Education	Configural invariance	2491.35 (261)	0.968	0.057 (0.054 - 0.059)	0.037	-	-	-	-
	Metric invariance	2558.42 (285)	0.968	0.055 (0.053 - 0.057)	0.039	67.07 (24)	-0.0006	-0.002	0.002
	Scalar invariance	2670.93 (309)	0.966	0.053 (0.052 - 0.055)	0.039	112.51 (24)	-0.0012	-0.001	0.001
	Residual invariance	3228.60 (339)	0.959	0.056 (0.055 - 0.058)	0.042	557.67 (30)	-0.0074	0.003	0.002

Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Root Mean Squared Error of Approximation (RMSEA), Standardized Root Mean Residual (SRMR), degrees of freedom (df). Δ signifies changes in a fit index value between different types of measurement invariance.

Before comparing the mean scores of the subsamples I examined the residuals and the distributions of responses to the scales. The residual versus fitted graphs showed that the residuals were approximately homoscedastic in all the subsamples (see Appendix Section 10.4.1). The boxplots indicated that the scales were generally left skewed, but distributions were generally similar between the different groups in the subsamples (see Appendix Section 10.4.2). One exception can be found in the health condition subsample, where the healthy public group showed relatively low levels of variance in the “Perceived Access to Options” scale compared to the other groups (see box-plot Figure 8 in the Appendix in Section 10.4.2). The variance of the healthy public group was 74.65, while the variance of the cancer group was 432.22 (the variance ratio between these groups is 5.79). Even though the assumption of equal variance across groups does not hold in this subsample, I still decided to conduct a one-way ANOVA and Tukey-Kramer test. This choice will be further elaborated in the discussion of this chapter.

Table 13 presents the mean scale scores of the groups in the different subsamples. The ANOVA of the health condition, age, and education subsamples indicated that the mean score of all the scales in at least one group differed significantly from the overall subsample mean in all subsamples (For all scales in all subsamples $p = < 0.001$). In terms of the t-test of the gender subsample, women scored slightly higher than men on the “Reflective Wellbeing” scale (difference of 0.51, $p = 0.25$), while men score slightly higher on the “Affective Wellbeing” scale (difference of 4.54, $p = <0.001$) and the “Perceived Access to Options” scale (difference of 2.51, $p = <0.001$).

The results of the Tukey-Kramer tests for the age group, education, and health conditions subsamples can be found in Table 14, Table 15, and Table 16 respectively. The scale scores generally followed a predictable pattern. With respect to education level, with an increasing level of education, individuals scored higher on each of the three scales of the WeRFree instrument. The differences between men and women were small. For the health condition subsample, the group of healthy participants had the highest scores on each of the three subscales. In contrast, particularly the participants affected by depression scored low on each of the three subscales. In terms of capabilities, the participants affected by arthritis also scored low, comparable to the participants with depression.

Surprising results could be identified in the age groups subsample. Perceived capabilities, measured with the subscale “Perceived Access to Options”, generally reduced with increasing age. However, the 65+ age group formed an expectation, since the reported level of capabilities was similar to participants in the age group of 35-44. The reflective wellbeing scale did not show much variation, except for the older age groups (55-64 and 65+), where the level of reflective wellbeing increased. A similar trend could be identified in the affective wellbeing group, where the level of affective wellbeing starts and keeps increasing from the 45-54 age group onward.

Table 13. Mean scale scores and associated standard deviation per subgroup per sample

Subsample	**Reflective wellbeing**		**Affective wellbeing**		**Perceived access to options**	
Age	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
18-24	62.01	19.01	63.61	20.23	87.51	14.82
25-34	62.57	19.50	64.04	19.46	84.90	16.89
35-44	61.51	20.78	64.32	20.64	80.63	19.94
45-54	61.68	20.62	66.46	21.08	77.07	21.22
55-64	65.18	19.78	70.71	20.05	76.92	21.31
65+	72.09	16.48	78.54	15.67	81.25	18.93
Gender						
Men	64.64	19.43	71.72	19.68	81.33	19.45
Women	65.15	20.08	67.18	20.40	78.82	20.44
Education						
High school	62.80	20.72	67.49	21.20	79.39	20.99
Some post-secondary, post-secondary certificate or diploma	64.93	19.49	69.52	20.03	79.40	20.01
University degree and higher	67.21	18.83	71.22	19.04	83.14	18.48
Health condition						
Healthy public	71.17	15.99	77.70	14.27	93.01	9.27
Arthritis	66.10	19.18	70.95	18.47	70.63	20.51
Asthma	65.34	18.24	68.77	17.98	82.13	18.41
Cancer	65.39	19.53	69.92	19.43	75.44	20.93
Depression	48.72	20.76	46.31	20.19	70.61	20.96
Diabetes	63.04	20.70	68.82	20.14	77.04	21.08
Hearing problems	68.72	18.21	74.10	16.65	85.05	15.54
Heart problems	65.74	18.98	71.60	19.64	76.59	21.09

Table 14. Tukey-Kramer test results for the age group subsample

Reflective wellbeing	18-24	25-34	35-44	45-54	55-64	65+
18-24	-	-	-	-	-	-
25-34	0.56†	-	-	-	-	-
35-44	-0.50	-1.05	-	-	-	-
45-54	-0.33	-0.89	0.17	-	-	-
55-64	3.17*	2.61**	3.67***	3.50***	-	-
65+	10.08***	9.52***	10.58***	10.41***	6.91***	-
Affective wellbeing						
18-24	-	-	-	-	-	-
25-34	0.43	-	-	-	-	-
35-44	0.71	0.28	-	-	-	-
45-54	2.85*	2.42*	2.14*	-	-	-
55-64	7.10***	6.67***	6.39***	4.25***	-	
65+	14.93***	14.50***	14.22***	12.08***	7.83***	-
Perceived access to options						
18-24	-	-	-	-	-	-
25-34	-2.62	-	-	-	-	-
35-44	-6.88 ***	-4.27***	-	-	-	-
45-54	-10.44***	-7.83***	-3.56***	-	-	-
55-64	-10.59***	-7.98***	-3.71***	-0.15	-	
65+	-6.26***	-3.64***	0.62	4.18***	4.33***	-

† Value should be interpreted as follows: the difference in reflective wellbeing score between age group scores, calculated as the score of age group 25-34 minus the score of age group 18-24. Other cells are to be interpreted similarly.

* p-value < 0.05

** p-value < 0.01

*** p-value <0.001

Table 15. Tukey-Kramer test results for the education subsample

Reflective wellbeing	High school	Some post-secondary, post-secondary certificate or diploma	University degree and higher
High school	-	-	-
Some post-secondary, post-secondary certificate or diploma	2.13***	-	-
University degree and higher	4.41***†	2.28***	-
Affective wellbeing			
High school	-	-	-
Some post-secondary, post-secondary certificate or diploma	2.03***	-	-
University degree and higher	3.72***	1.69**	-
Perceived access to options			
High school	-	-	-
Some post-secondary, post-secondary certificate or diploma	1.34*	-	-
University degree and higher	5.08***	3.74***	-

† Value should be interpreted as follows: the difference in reflective wellbeing score between groups that differ in terms of level of education, calculated as the score the group with a university degree or higher as the highest level of education minus the score of the group with high school as the highest level of education. Other cells are to be interpreted similarly.

* p-value < 0.05

** p-value < 0.01

*** p-value <0.001

Table 16. Tukey-Kramer test results for the health condition subsample.

	Healthy public	Asthma	Cancer	Depression	Diabetes	Hearing problems	Arthritis	Heart problems
Reflective wellbeing								
Healthy public	-	-	-	-	-	-	-	-
Asthma	-5.83***†	-	-	-	-	-	-	-
Cancer	-5.79***	0.05	-	-	-	-	-	-
Depression	-22.46***	-16.62***	-16.67***	-	-	-	-	-
Diabetes	-8.13***	-2.30	-2.35	14.32***	-	-	-	-
Hearing problems	-2.45*	3.38**	3.34**	20.01***	5.68***	-	-	-
Arthritis	-5.07***	0.76	0.71	17.38***	3.06*	-2.62	-	-
Heart problems	-5.43***	0.40	0.36	17.03***	2.70*	-2.98*	-0.36	-
Affective wellbeing								
Healthy public	-	-	-	-	-	-	-	-
Asthma	-8.93***	-	-	-	-	-	-	-
Cancer	-7.79***	1.14	-	-	-	-	-	-
Depression	-31.40***	-22.47***	-23.61***	-	-	-	-	-
Diabetes	-8.88***	0.05	-1.09	22.52***	-	-	-	-
Hearing problems	-3.60***	5.33***	4.18***	27.79***	5.27***	-	-	-
Arthritis	-6.75***	2.18	1.03	24.64***	2.13	-3.15**	-	
Heart problems	-6.10***	2.82*	1.68	25.29***	2.78*	-2.50	0.65	-
Perceived access to options								
Healthy public	-	-	-	-	-	-	-	-
Asthma	-10.88***	-	-	-	-	-	-	-
Cancer	-17.57***	-6.68***	-	-	-	-	-	-
Depression	-22.40***	-11.52***	-4.84***	-	-	-	-	-
Diabetes	-15.97***	-5.09***	1.60	6.43***	-	-	-	-
Hearing problems	-7.96***	2.92**	9.61***	14.44***	8.01 ***	-	-	-
Arthritis	-22.38***	-11.49***	-4.81***	0.03	-6.41***	-14.14***	-	-
Heart problems	-16.42***	-5.54***	1.15	5.98***	-0.45	-8.46***	5.96***	-

† Value should be interpreted as follows: the difference in reflective wellbeing score between groups that differ by disease, calculated as the score of the group affected by asthma minus the score of the healthy group. Other cells are to be interpreted similarly.

* p-value < 0.05

** p-value < 0.01

*** p-value <0.001

5.4 Discussion

In this chapter, I tested the measurement invariance properties of the WeRFree instrument. The instrument showed to have configural, metric, and scalar invariant properties in the tested subsamples. Full measurement invariance was established in the subsamples where individuals were grouped according to their gender or education. The establishment of scalar invariance in every sample indicates that instrument scores of the WeRFree instrument can be compared across different groups.

The mean differences in instrument scores between different groups were generally unsurprising. In the education groups sample, a pattern could also be identified, in which higher levels of education were associated with improved subjective wellbeing and improved health-related capabilities. This result is generally in line with the literature, where education is linked to higher levels of subjective wellbeing when controlled for other socioeconomic variables (such as income or health) (Dolan et al., 2008). Similarly, the association between health and education has been well documented (Cutler and Lleras-Muney, 2006). It is therefore not surprising that an increase in education level is associated with higher mean scores in the “Perceived Access to Options” scale.

Regarding the health condition subsample, the relatively unequal level of variance in the healthy condition sample could be considered a problem with a one-way ANOVA and the Tukey-Kramer test, particularly because of the differences in the sample size between the different groups. Regarding the one-way ANOVA, p-values should be interpreted with care, since the larger healthy public subgroup showed lower levels of variance than other subgroups. Under these circumstances, the ANOVA is too liberal in terms of Type I error (Blanca et al., 2018). In order to have some indication of whether there are significant differences in the samples, the decision was still made to apply the ANOVA. Its results should however be interpreted exploratory and interpreted with care. Similarly, the results of the Tukey-Kramer test should be interpreted exploratively. Nevertheless, the results may be robust, as a study by Ramsey et al. (2010) indicates that the Tukey-Kramer test is robust as long as the variance ratio is less than 8 in the groups that showed the largest differences in variances. This is the case in the study presented in this chapter.

When comparing the means in the health subsample, it should be noted that the groups affected by disease scored generally lower on the “Reflective Wellbeing”, “Affective Wellbeing” and “Perceived Access to Options” scales when compared to healthy individuals. Particularly participants with depression scored significantly lower on all scales of the WeRFree instrument. Unsurprisingly, individuals affected by depression score the lowest on the “Affective Wellbeing” and “Reflective Wellbeing” scale, given the overlap between the experience of depression and the experience of subjective wellbeing (Luhmann et al., 2012). Also, the comparatively low score of depressed individuals in the “Perceived Access to Options” scale, which is as low as the score of individuals affected by arthritis, has been observed in the literature. For example, a study by Gaynes et al. (2002) showed a similar effect of arthritis and depression on role functioning and physical functioning, which are variables that resemble the items in the “Perceived Access to Options” scale. It thus seems that the “Perceived Access to Options” scale is able to measure the effects of physical and mental health on capabilities.

In the age groups subsample, participants in the older age groups (55-64 and 65+), scored higher in the reflective – and affective wellbeing scales compared to the other age groups, which is a result that is supported by the literature (López Ulloa et al., 2013). One unexpected result could however be observed. The 65+ age group broke a pattern of diminishing health-related capability with increasing age. One potential explanation for this observation is related to the limitations of the MIC survey database, which will be discussed in the next section.

5.4.1 Limitations

The recruitment strategy of the MIC study aimed at recruiting a sufficient number of participants from different health backgrounds for their database (Richardson et al., 2012). As such, the database was not necessarily designed to reflect specific (sub-) populations. This might be an explanation for why the participants that are older than 65 had level of capability that was comparable to participants in the 35-44 year age group. The results of the scale score comparison should therefore only be generalized with care and should at best be used to form hypotheses for further studies.

Another limitation concerns the use of the MIC study database to both develop an instrument and test the measurement invariance properties of that instrument. Consequently, measurement errors that can be attributed to the design of the MIC survey may be unaccounted for due to using the same database for both of these studies. Therefore, the measurement models might overfit, which in the context of the study presented in this chapter means that the measurement invariance properties of the WeRFree instrument can be overestimated.

5.4.2 Implications

Essentially, after measurement invariance testing, there are two possible outcomes: either an instrument is or is not measurement invariant. Establishing measurement invariance indicates that individuals interpret the items of instruments similarly. By comparing the responses of disadvantaged individuals (e.g., with a disease) with a reference group (e.g., healthy individuals), one can establish whether responses are affected by adapted preferences. Although not routinely, reference groups of healthy individuals have already been used to test for response shift in patient groups (Sajobi et al., 2018; Schwartz et al., 2006).

The identification of such a reference group might be a challenge since it should be a group with an extensive set of (health-related) capabilities to ensure that adapted preferences do not affect the responses of this reference group. However, what such a set entails is not entirely clear (see the discussion around capability lists in section 1.5.2), which complicates the identification of a reference group. More research is thus necessary. However, for the time being, in the context of testing for adapted preferences in individuals affected by health issues, it might be sufficient to use a sample from a reasonably healthy population.

When measurement invariance cannot be established, further studies can be conducted to identify the source of measurement noninvariance (Jung and Yoon, 2016; Putnick and Bornstein, 2016). In this case, the source of measurement noninvariance does not necessarily have to be adapted preferences, since there can be several alternative explanations for why individuals interpret items differently. Depending on the extent of measurement noninvariance and its nature (for example configural, scalar, metric, or residual), researchers need to decide whether measurement noninvariance poses a problem for the interpretation and comparison of instrument

scores. If problematic measurement noninvariance is identified, researchers need to study whether it is possible to account for this noninvariance, or conclude that a direct comparison of scores is not possible (Putnick and Bornstein, 2016).

Within the capability approach and conventional health economics, patient self-reports are sometimes viewed skeptically because of adapted preferences. In the context of this discussion, it should be noted that if measurement noninvariance is identified and the noninvariance can be attributed to adapted preferences, researchers should still be careful about dismissing information from disadvantaged individuals, as these experiences are still informative (Khader, 2013; Robeyns, 2017b). In the context of wellbeing and health, this notion might be particularly relevant. It could be argued that in the face of disability and ill health (in particular chronic illness), individuals make a conscious and intentional effort to change their preferences toward the things that they are still able to do (Mitchell, 2018). Individuals might even fundamentally change their ideas about what it means to be healthy, from a view of health as representing high physical functioning to a holistic view of health in which it is important insofar as it supports or limits living a valuable life (Mitchell, 2018). In this context, self-reported information by individuals give a richer understanding of their wellbeing.

5.4.3 Conclusion

To summarize, the results presented in this chapter should be interpreted as the first evidence for the measurement invariance properties of the WeRFree instrument in the various subsamples of the MIC study. This also provides the first evidence that responses to the items in the WeRFree instruments were not affected by adapted preferences. However, due to the limitations of this study, the results need to be confirmed in different samples. In the context of the development of capability instruments, future studies should focus on establishing the measurement invariance properties of instruments used in health economics. Information from scales can then be used to assess the effect of health technologies and inform value assessment.

6 DISCUSSION

This dissertation studies the use of the capability approach in instrument development for HTA. A major part of this dissertation concerns the development of the WeRFree instrument, which is a capability instrument that is based on a theoretical framework that was developed with the concept of option freedom. I have argued that this instrument could assess wellbeing comprehensively with its three scales, which cover both health-related capabilities and subjective wellbeing.

It should be noted that the WeRFree instrument is still in an early stage of development. Therefore, the work presented in this dissertation warrants further discussion. In Section 6.1, the discussion covers the general limitations of the work presented in this dissertation in relation to instrument development. In Section 6.2, I will reflect on some theoretical issues regarding the information that the WeRFree instrument provides. In this section, I will discuss the use of information about subjective wellbeing in the context of the capability approach, whether the WeRFree instrument is a sufficient source of information in the context of HTA, and how information from the three different scales could be integrated. In Section 6.3, the development of the WeRFree instrument will be discussed in light of the criteria for capability list development of Robeyns (2005). In Section 6.4, the most important future research needs that are discussed in this chapter are summarized.

6.1 Limitations

The development of the WeRFree instrument has been a fairly theoretical exercise. Although it is an important step in the development of a capability instrument that an initial idea is developed of what should be measured, it should also be acknowledged that in its current state, the instrument lacks validity. The content validity of the WeRFree instrument will need to be further validated together with experts and users of the instrument. Due to the available resources, it was not possible to conduct such interviews within the scope of the project presented in this dissertation. Still, it is important to note that the instrument's content validity will need to be studied before the WeRFree instrument is used in practice. Additionally, further criterion – and construct validity tests will need to be conducted to study whether the constructs as measured by the WeRFree instrument are correlated with other variables as hypothesized. It was not possible to conduct these tests with the MIC database, since

the items that were used in the development of the WeRFree instrument come from instruments that could be used for such a study. As such, new studies will need to be conducted to establish criterion – and construct validity.

Evidence regarding the reliability of the WeRFree instrument is also lacking. The reliability of the WeRFree instrument in terms of Cronbach's alpha will need to be established in new samples. Furthermore, its test-retest reliability has not been established. As such, the instrument will need to be applied in fresh samples to establish whether the instrument is reliable in different contexts. Lastly, the MIC study, from which the data were used to develop the WeRFree instrument, was not constructed to represent any type of population. As such, evidence from this study (such as mean scores, or measurement invariance properties) should be generalized with care.

Beyond limitations related to the reliability and validity of the WeRFree instrument, one further limitation is that the WeRFree instrument consists of three scales. This complicates assessments since it is unclear what the relative importance of each scale is in relation to overall wellbeing. In the context of HTA, this means that this complicates the assessment of the value of health technologies that vary in their effects on the different scales. It is therefore important to conduct further research on how to integrate information from these scales. Possible methods of doing so are further discussed in Section 6.2.3.

6.2 Reflections on the information provided by the WeRFree instrument

6.2.1 Subjective wellbeing in capability instruments

The WeRFree instrument consists of three scales: "Reflective Wellbeing", "Affective Wellbeing", and "Perceived Access to Options". The reflective – and affective wellbeing scales reflect subjective wellbeing. From a measurement perspective, the moderate correlations indicate that the constructs reflect separate constructs (see Chapter 4). This implies that these three scales provide supplemental information for wellbeing assessment. From a theoretical perspective, there are however some complications. Sen has extensively critiqued the use of subjective wellbeing information (see Section 1.4). Instead, Sen has argued (particularly in his later writings) that using information from capabilities is in theory *sufficient* to assess wellbeing (Clark, 2005). In the context

of capability instrument development, this would mean that instruments should consist of content that focuses on the measurement of capability. These theoretical considerations have resulted in the development of capability instruments that solely focus on assessing wellbeing in terms of capabilities (see Chapter 2). This raises the question of whether scales that reflect subjective wellbeing should be part of capability instruments, or, in the context of the WeRFree instrument, whether the affective and reflective wellbeing scales should remain part of an instrument that is based on the capability approach.

Before reflecting on this question, it is helpful to provide some context to the development of the capability approach itself. Sen developed the capability approach as an alternative evaluative framework in light of the limitations of the welfarism, which uses information about individual utilities as the basis of wellbeing assessment. In this light, it is unsurprising that much of Sen's work evolves around the benefits of assessing wellbeing in terms of capabilities and the limitations of information about subjective wellbeing (Clark, 2005). However, this focus also results in ambiguities in Sen's theory, since Sen never precisely discussed how capability and subjective wellbeing are related (Clark, 2005).

When Sen discusses the relationship between capability and subjective wellbeing, he uses two examples: the act of cycling (Sen, 1997) and the act of eating bread or rice (Clark, 2005; Sen, 1984). Both these examples essentially follow the same pattern: there is a commodity (bike or bread), which has certain characteristics (transportation or providing nutrition (Clark, 2005; Sen, 1984, 1997)). These characteristics can be converted into functionings (such as cycling, or being well-fed). In turn, these functionings result in some level of "utility", which in this context should be understood as a "mental state". What this mental state exactly entails is however not clear (Clark, 2005). It is also unclear how to include differences in how people experience their capabilities in wellbeing assessment, since according to Sen the assessment of capabilities is sufficient in such contexts (Clark, 2005; Fleurbaey, 2006). However, two individuals with the same level of capability could have a very different life, with associated differences in how they experience their wellbeing (Fleurbaey, 2006). Clark (2005) has therefore argued that the role of subjective experiences should be expanded since measuring these would bring additional information that can be used for the assessment of wellbeing.

The inclusion of information about how individuals experience their capabilities might even lead to an increased understanding of the value of those capabilities themselves. In the context of option freedom, it could be argued that low levels of subjective wellbeing combined with high levels of capability is an indication that an individual is not content with the options available to her or him, or, in other words, that the options do not have much subjective significance to this individual. Consequently, information about the reflective and affective wellbeing of individuals gives information about the subjective significance of those options to individuals (see Section 1.5.1 for a further explanation of subjective significance in the context of option freedom). There are thus empirical and theoretical arguments for assessing wellbeing in terms of both capabilities and functionings, in particular functionings related to subjective wellbeing. The WeRFree can therefore be seen as a proposal for a more comprehensive measure of capability wellbeing.

6.2.2 Sufficiency of the information provided by the WeRFree instrument

In the context of wellbeing assessment, the question is what kind of information is provided by the WeRFree instrument and perhaps more importantly, what kind of information might be missing. In this respect, a comparison of the content of the WeRFree instrument with other capability instruments that have been developed to assess wellbeing across different disease groups and different contexts may provide further insights. These instruments are the following: the ICECAP-A, ICECAP-O, CALY, and OCAP-18. From these instruments, a comparison cannot be made with the CALY. The CALY is at the time of writing this dissertation still under development and a list of items has not yet been published.

Between these instruments, the most direct comparison of the WeRFree instrument can be made with the ICECAP-A, since three items of the ICECAP-A were part of the measurement model that was used to test the theoretical framework from Chapter 3. These three items loaded with reasonable strength on the different factors of the measurement model (see Table 6). Engel et al. (2017) also conducted an exploratory factor analysis with the ICECAP-A and a set of HRQoL instruments (15D, AQoL-8D, EQ-5D-5L, Health Utilities Index Mark 3 (HUI-3), and Short-Form Six-Dimension (SF-6D)). Engel et al. show that in such a factor analysis, the items of the ICECAP-A primarily load on one factor that broadly covers emotional wellbeing. Further (cross-)

loadings of single items can be found on factors covering social wellbeing or factors covering physical health-related aspects. Based on insights from this study and the publication by Engel et al., it can be concluded that the WeRFree instrument probably covers the content of the ICECAP-A reasonably well.

The content of the ICECAP-O and the ICECAP-A show large similarities (Al-Janabi et al., 2012). These similarities suggest that is probable that the WeRFree instrument covers the content of the ICECAP-O as well. Less overlap is however to be expected between the WeRFree and the OCAP-18, given that the latter instrument includes items related to enjoying the environment, or the ability to express views, which the WeRFree instrument is missing.

What sets the WeRFree instrument apart from the three other capability instruments is that the WeRFree instrument contains more items that reflect health-related capabilities due to its “Perceived Access to Options” scale. These items come from existing HRQoL instruments. Studies comparing the ICECAP-A and ICECAP-O with HRQoL instruments indicate that such health-related items provide supplementary information to the ICECAP-A and ICECAP-O (Davis et al., 2013; Engel et al., 2017; Franklin et al., 2018; Helter et al., 2020; Keeley et al., 2016). Given that the OCAP-18 includes only two items related to health, it is probable that the WeRFree instrument is more sensitive to health-related capability as well.

It should however be noted that the comparisons of the WeRFree instrument with the OCAP-18, ICECAP-A, and ICECAP-O are based on either a factor analysis with a limited number of items or a comparison of the content of the instruments. Consequently, these comparisons are limited and the observations in this discussion should therefore be understood as hypotheses that can be the subject of future studies.

Beyond comparisons with other capability instruments that have been developed for similar use, it should also be noted that the content of the WeRFree instrument is not disease-specific. It is therefore probable that the instrument is less or even insensitive to important domains of wellbeing that are relevant for specific patient groups. This point is illustrated by comparing the content of the WeRFree instrument with the content of the Abbreviated Profile of Hearing Aid Benefit (APHAB) (Cox and Alexander, 1995). The APHAB is a disease-specific instrument that is used to measure HRQoL in

people affected by hearing loss. The instrument consists of 24 items that cover different aspects of hearing loss, such as the ability to follow conversations, or the ability to listen to a lecture. Such disease-specific items are missing on the WeRFree instrument.

Consequently, when the WeRFree instrument is used to assess wellbeing, it is important to consider if additional information is needed. This can be in the form of alternative instruments that cover disease-specific elements of wellbeing that are otherwise not represented in the WeRFree instrument. In the context of cancer research, this approach is already followed. The functional assessment of cancer therapy-general (FACT-G) instrument is developed as a general instrument to assess HRQoL in cancer patients. In addition to the items of the general instrument, researchers have also developed additional cancer-type-specific scales to capture additional elements of cancer-type-specific HRQoL that are not well reflected in the general instrument (Victorson et al., 2008). In the context of the WeRFree instrument, it may be an interesting idea to complement information from the instrument with a disease-specific instrument to fully assess the wellbeing of individuals.

A follow-up question would be how to identify elements of wellbeing that need to be part of an informational base for value assessment but are not captured by the WeRFree instrument. One way of identifying such information would be to conduct a systematic review that aims to identify what kind of elements of wellbeing are important for specific patient populations. Furthermore, content from instruments that have been developed to assess HRQoL or wellbeing in these specific populations could be a source of information about potentially relevant elements of value.
Additionally, one can also consider conducting qualitative studies within the patient population to evaluate what is important to their wellbeing (in terms of capabilities and subjective wellbeing) and how a new health technology changes this experience positively or negatively. Such a study has been conducted by the Danish Health and Technology Assessment Council (Monitorering & Medicinsk Teknologivurdering, 2009), in which additional effects of a health technology on cancer patients were identified that were not captured by the instruments that were part of their study.

The selection of such elements of wellbeing with qualitative research can however be affected by adaptive preferences (Robeyns, 2017b). Thus, qualitative research that is

conducted to identify these elements may need to use specific methods to identify elements that are affected by adaptive preferences. In this context, Khader (2011) proposes the “deliberative perfectionist approach” to identify whether preferences are adapted. In this approach, researchers conduct a qualitative study to understand how people experience their wellbeing. This study should follow a deliberative approach. While deliberating with the group under research, researchers keep a basic idea of flourishing in mind and study whether any deviations from this idea are a result of adaptive preferences or just part of rationally adapted preferences. What this flourishing exactly entails would need to be further specified, but a start could be to adapt existing capability lists to the German context with the methods explained in Section 1.6.2, in which the development of capability lists is explained.

6.2.3 Weighing of scales

As was mentioned in the overall limitations in Section 6.1, one challenge in the use of the WeRFree instrument is that it consists of three different scales. Information from these scales needs to be integrated to come to an overall wellbeing assessment of (groups of) individuals to facilitate inter- and intrapersonal comparisons. One way to integrate information from the different scales is by developing weights, similarly as is done for the QALY.

The development of weights is however a relatively underdeveloped topic in the capability approach. Robeyns (2006) observed that compared to the literature written on dimension selection and list development, little literature is written about how to weigh different dimensions. Furthermore, the authors that have written about this subject recommend very different approaches in the context of the capability approach (Robeyns, 2017d). For instance, Nussbaum (2003) argues that the ten capabilities on her list should be guaranteed to a certain level for every citizen. Prioritization should only temporarily be applied when strictly necessary. Alternatively, in the context of health economics, weights have been developed for various capability instruments to facilitate decision-making (Helter et al., 2020).

With this discussion in mind, the question is whether weights should be developed in the first place for the WeRFree instrument. In this, I follow the view of Robeyns that not developing weights might be defendable from a philosophical perspective, but could be difficult to justify in practice. This is because resources are limited, so choices

regarding resource allocation have to be made, and any such choice results in an implicit or explicit weighing of priorities (Robeyns, 2006). The development of explicit weights at least results in transparency and the methodology for developing the weights can be publically scrutinized. What kind of method should then be used for weight development?

6.2.3.1 Methods that can be used for developing weights

Decancq and Lugo (2013) identified different methods to weigh dimensions that can be categorized into three groups:

1. Data-driven methods to estimate weighs.
2. Normative methods to develop weights.
3. Hybrid methods to develop weights.

Data-driven methods use descriptive data to establish weights. A practical example of this method is frequency-based weights. These weights are based on the distribution of a population in a certain domain. To illustrate, one can establish frequency-based weights to establish shortfalls related to the accessibility of certain types of food. In this context, one can establish that only a small proportion of the population eats caviar, so being unable to eat caviar seems to be unimportant and does not require to have a high weight. Alternatively, a large proportion of the population consumes bread, which signifies its importance as a source of food in a population. Consequentially, a higher weight should be given to the ability to eat bread, which illustrates the shortfall that the smaller part of the population experiences by not being able to do so.

Normative-driven weights are derived from value judgements about the trade-offs between domains. An example of this method is eliciting weights from experts. Hybrid approaches aim to combine information about the distribution of achievement in domains with valuations by individuals. One example is the stated preference method, which has also been discussed in Section 1.3.2.

How applicable are these methods generally for the capability approach? As Robeyns (2017d) mentions, this depends on the research question and the context in which the research takes place. The capability approach is flexible enough to incorporate any of these methods, as long as researchers are explicit about the implied value judgements

that are associated with the use of each of these methods. Therefore, the question is how applicable each of these methods is in the context of the WeRFree instrument.

In this context, one limitation of the data-driven approaches is that they are essentially based on descriptive methodology. To illustrate, it might be possible that frequencies from frequency-based rates do not reflect the importance that a population attaches to a certain dimension (Decancq and Lugo, 2013). For instance, in the scenario presented above, it might be the case that people who do not eat bread also do not want to eat bread, but eat something else instead. This means that individuals that do not eat bread might in fact not be worse off than individuals who eat bread (Decancq and Lugo, 2013). The alternative would then be to use normative approaches to set weights, such as the elicitation of weights from experts. However, weights set by experts could be argued to be paternalistic (Decancq and Lugo, 2013). It is also unclear which experts should be asked for such a valuation exercise (Decancq and Lugo, 2013). Given the limitations of both normative- and data-driven approaches, it might be best to follow a hybrid approach to elicit weights. A properly constructed hybrid approach might result in estimates that are not overly paternalistic since its weights are elicited from the general population. Furthermore, these weights reflect normative value judgements that are grounded in what the population values (Coast et al., 2008a; Cookson, 2005).

6.2.3.2 One set of fixed weights or flexible context-dependent weights

Another discussion is whether one fixed set of weights should be developed that can be used in different contexts (e.g., different patient groups) or if a flexible set of weights should be developed (e.g., for every health technology assessment a unique set of weights). In the context of the capability approach, Sen has argued against the use of fixed weights that can generally be applied in different contexts (Sen, 2004). As an example, Sen noted that the relative value of being nourished is higher than the value of having adequate shelter for an individual with a home living in an area struck by famine. Vice-versa, to an individual without a home but with adequate access to food the relative value of being sheltered will be higher than having food (Sen, 2004). In practice, the ICECAP series of instruments can be seen as reflecting these ideas, by developing different instruments for specific groups with associated tailored weights (Coast, 2019).

Additionally, there are also empirical concerns with the use of fixed weights that are used generally in different populations. Richardson et al. (2015) conducted a study where the effect of weighing was disentangled from instrument score ranges and instrument descriptive systems. In this context, the score range reflects the minimum possible utility value to one (e.g., for certain tariffs of the EQ-5D-5L the utility range goes from -0.51, a utility state worse than death to 1, which represents the utility of perfect health) and the descriptive systems concern the actual content of the instruments. In their study, Richardson, Iezzi, and Khan found that general (utility) weight adjustments to instrument scores generally increase the differences in scores between multi-attribute utility instruments, when disentangling the effect of weighting from the score range and descriptive systems.

This result is counter-intuitive, since these weights are theoretically supposed to reflect the same construct, utility, and one would expect that weight adjustments to raw instrument scores would result in decreasing the differences in utility-adjusted scores of different instruments in groups of people with varying health conditions (Richardson and Iezzi, 2014). To illustrate, take two hypothetical HRQoL instruments. Both instruments consist of three items. Instrument A has one item covering physical health aspects and two items covering mental health aspects of HRQoL. Instrument B is an instrument with two items covering physical health aspects and only one item covering mental health aspects of HRQoL. Due to these differences, instrument A is more sensitive to physical health and instrument B is more sensitive to mental health. Based on raw scores, it is possible that an individual with impaired mobility scores lower on instrument A than on instrument B. When adjusting the raw scores of these instruments with weights, it is possible to account for this difference, since under-represented items will theoretically receive higher weights and over-represented items receive lower weights. To further illustrate, the utility derived from being mentally healthy can be allocated to only one item, and the utility derived from being physically healthy to two items in instrument A. This means, that the mental health item becomes relatively more important and the instrument becomes relatively more sensitive to changes in mental health. A similar set of weights for the items of instrument B (which makes the physical health item relatively more important compared to the mental health items) would theoretically result in the instrument scores being more comparable (Richardson and Iezzi, 2014; Richardson et al., 2015).

One possible explanation for the finding of Richardson and Iezzi (2014) is that these general weights are the product of algorithms that overfit a particular dataset, which reduces the validity of the weight adjustments (Richardson and Iezzi, 2014). Consequently, Richardson and Iezzi argue that general weights that are based on intricate algorithms should not be used to adjust the scores of instruments. Instead, they argue for using a simple adjustment to the score range with help from a third variable.

Based on these empirical and theoretical considerations, I would therefore argue for developing a flexible set of weights that reflect the importance of its scales in different contexts. This third variable could be the length of life derived from a TTO, on which each of the minimum – and maximum scores of the three scales can be rescaled. Such a set of weights should ensure that the effects of health technologies are appropriately valued.

6.2.3.3 Who should be the source of weights?

A related question is from which population flexible weights should be elicited with a hybrid approach. In the context of health economics, a general distinction is made between eliciting weights from a sample of the general population or a sample of a patient population (Versteegh and Brouwer, 2016). In practice, weights from the general population are typically used to inform health technology assessment. These weights are applied in different settings and different patient populations to calculate QALYs (Section 1.3.2 for an explanation of how QALYs are calculated). From a normative perspective, both sources of weights have their merits and it is difficult to say which samples should be used from a theoretical perspective (see for an overview of the discussion Versteegh and Brouwer (2016)). It could be argued that weights from patients are unreliable due to adaptive preferences. Additionally, one could argue that weights from the general population reflect the value of “insuring” oneself against disease. These arguments favor the use of generally applicable weights elicited from the general population. Alternatively, it could be argued that patients are the actual experts that should be involved in weight elicitation since they have the lived experience of being in a certain health state. Consequently, the discussion has not been solved in the general health economics literature and it is unclear what kind of sample should be used to elicit weights for the WeRFree instrument.

Still, measurement invariance testing between a patient sample and a sample of the general population might be helpful before eliciting weights. The establishment of measurement invariance indicates that people in these samples evaluate their capabilities and subjective wellbeing similarly in a specific instrument (see Chapter 5). If these groups evaluate their own capabilities and subjective wellbeing similarly, one could argue that the preferences of the patient groups are as least as well informed as that of the general population concerning the constructs and items of the WeRFree instrument. If measurement invariance is established, then one could argue that the preferences of patients are at least as informed as the preferences of the general population with respect to the weighted items. If measurement invariance is not established, it could point to potential weaknesses in the weighing exercise, since the informational inputs are interpreted differently and weights are developed for incomparable constructs. In such circumstances, additional qualitative research can be conducted to study how individuals weigh different scales of an instrument.

To conclude this section, for certain contexts the weighing of domains is useful to answer research questions or inform policy-making. In the context of using multi-dimensional or multi-scale self-report instruments, insights from psychological and health economic research indicate that general weights do not accurately reflect the relative value of an instrument's different attributes. Instead, a flexible set of weights needs to be developed. These weights should reflect the relative importance of each of the WeRFree's scales in the context of a certain health technology. The specific methodology of how to elicit weights can be subject to future research.

6.3 The WeRFree instrument and criteria for developing capability lists

A major part of this dissertation shows the first stages of the development of a capability instrument. The development of such an instrument essentially means that a capability list has been developed. As such, it is important to reflect on the current state of the WeRFree instrument in the context of capability list development. To recall, Robeyns (2005) developed four criteria for the development of such lists. These four criteria are:

Criterion (1): A capability list should be explicitly formulated, discussed, and defended.
Criterion (2): The methods for developing the list should be justified.

Criterion (3): Two versions of a capability list should exist: an “ideal” and a version that can be applied in the constraints of empirical research.
Criterion (4): The capability list should be exhaustive and non-reductive.

Concerning criterion (1), it should be noted that the formulation of the content of the WeRFree instrument is presented in Chapter 3 and Chapter 4. In terms of defending the content of the WeRFree instrument, the most direct comparisons can be made with other capability instruments. Such a comparison is provided in Section 6.2.2. The conclusion of this section is that based on the similarities and differences in the content between the WeRFree instrument and the other capability instruments, we expect the WeRFree instrument to be more sensitive to how health affects the capabilities of individuals while being equally sensitive to changes in subjective wellbeing. It should however be noted that particularly the OCAP-18 has more content that is relevant for the measurement of changes in capabilities through social and policy interventions. The WeRFree instrument is thus not the most appropriate instrument to assess capability wellbeing in every context due to its focus on health.

Regarding criterion (2), I developed the WeRFree instrument in four steps. The first step was to broadly and critically read about the capability approach, wellbeing assessment, and the combination of both fields applied in the field of health. This formed the basis for the analysis of Chapter 2, in which I concluded that the concept of capability by Sen might obstruct authors from recognizing burdens in capability achievement. Furthermore, Sen’s concept of capability could impede authors’ in their decision making whether certain elements of wellbeing should be assessed in terms of capability or with alternative concepts that would yield more or additional information about wellbeing. This step thus gave me a broad overview of the literature and helped me identify problems in the current uses of the capability concept in wellbeing assessment in the field of health.

With the learnings from the first step in mind, the second step was to apply the concept of option freedom in a synthesis of qualitative papers that support the development of various capability instruments. This synthesis, presented in Chapter 3, ensured that the capability lists that were presented in the different qualitative papers were integrated into one framework. As such, the synthesis is rooted in the observations of researchers about what kind of elements of wellbeing are important to individuals,

which means that the elements of these capability lists are reflected in the resulting theoretical framework.

In the third step, which is presented in Chapter 4, such an instrument was developed using psychometric methods. This step supports the development of a capability list by helping understand the dimensionality of the operationalized constructs and identifying the items that yield the most information to measure the different elements of the theoretical framework developed in Chapter 3. The result is that the scales of the WeRFree instrument are informative and that the scale scores reflect single constructs (i.e., the scales are unidimensional). Reflections on whether subjective wellbeing constructs should be part of a capability instrument as well as the sufficiency of the informational value of the WeRFree instrument are the subject of Section 6.2.1 and Section 6.2.2.

The fourth step is to provide the opportunity to the public to debate the content of the WeRFree instrument in the context of it being a capability list. This will take the form of scientific publications which present the different stages of instrument development that were addressed in Chapter 3 and Chapter 4. Scientific publications are however not an ideal method of presenting and allowing to debate the content of the WeRFree instrument, since it is probable that only a small and unrepresentative part of the public (scientific authors who have an interest in the capability approach) could potentially comment on the list. Further research involving patients and members of the public is therefore necessary.

In relation to criterion (3), it should be noted that in this dissertation an ideal list and an empirical list have been developed. The ideal list of capabilities is the product of the best-fit framework synthesis in Chapter 3. Essentially, this list consists of the subthemes of this framework, which show single elements that are important for the assessment of health-related capability wellbeing according to my synthesis of previous qualitative research. As mentioned, the content of the WeRFree instrument does not reflect all the subthemes of the qualitative framework developed in Chapter 3. As such, the content of the WeRFree instrument that was developed in Chapter 4 represents the applied version of this list. Future research should therefore establish whether the WeRFree instrument should be adjusted to reflect these missing elements.

Regarding this research question, studying the content validity of the WeRFree instrument might add additional insights.

Some further observations can be made with respect to criterion (4). It should be noted that for the WeRFree instrument, the decision was made to not include a scale that reflects the construct “Perceived Control”. The exclusion of the “Perceived Control” domain for the instrument could be considered at odds with the importance of an exhaustive and non-reductive list. However, the empirical evidence shows that most of the information from the “Perceived Control” construct is captured by the “Affective Wellbeing” construct since both domains show a correlation coefficient that is higher than 0.9. When considering the trade-off between not burdening patients with instruments that are too long and instruments that are comprehensive enough to capture a sufficient amount of information for evaluative purposes, I decided that the “Perceived Control” construct can be excluded from the WeRFree instrument without it having too large of an impact on the information that it provides.

To conclude this section, it should be clear that there are advantages and limitations in the research conducted in this dissertation in relation to the development of capability lists. As was mentioned, one of the weaknesses of the WeRFree instrument is that its development has mostly been theoretical. However, this weakness has been beneficial since it forced me, the developer, to reflect on the theory of the capability approach, which led to the use of the concept of option freedom for instrument development. This facilitated the identification of relevant elements of capability that could have been missed with Sen’s definition of capability, such as the burdens that people experience in capability achievement. Lastly, the synthesis of qualitative literature helped the integration of various views of what is important for individuals. As such, it is unsurprising that the content of the WeRFree instrument reflects elements that are important in both capability and psychological research. Even though of these advantages, it should be clear that further efforts are needed to justify the value judgements that are the foundation of the content of the WeRFree instrument. These will be summarized in the next section.

6.4 Future research

More research is thus needed before the WeRFree instrument can be used in practice. The immediate next step is to further validate and justify the content validity of the

WeRFree instrument with its users. One method for doing so is by conducting studies, such as focus group discussions, with groups of individuals that vary in their health condition. These studies should aim to establish whether the content of the WeRFree instrument covers all the elements that are important for the assessment of their (health-related) capability wellbeing in different populations. Such studies have for example been conducted for the ICECAP-O to establish whether the instrument sufficiently covers all the relevant capabilities of a group of individuals (Hörder et al., 2016; van Leeuwen et al., 2015). Coincidently, these studies and the potential changes that might occur to the WeRFree instrument as a consequence of these studies will strengthen the justification of the content of the WeRFree instrument. Moreover, these studies will highlight the limitations of the WeRFree instrument with respect to measuring wellbeing in various populations and indicate whether the instrument needs further adaptation.

After establishing content validity and possibly further improving the WeRFree instrument, its criterion and construct validity should be studied in different settings. For this, the (updated) WeRFree instrument will need to be used alongside other HRQoL instruments, subjective wellbeing instruments, and capability instruments to study whether hypothesized relationships can be established in various populations. In the context of establishing construct validity, further cognitive pretests could additionally be conducted to understand if the items of the WeRFree instrument measure perceived capabilities, operationalized as “Perceived Access to Options”, and experienced wellbeing, operationalized in the scales “Reflective Wellbeing” and “Affective Wellbeing” (Koskey, 2016). Examples of cognitive pretests and a focus group discussion can be found in Section 10.5 of the Appendix. Combined with quantitative data that explore how the scales of the WeRFree instrument relate to other variables, this could result in sufficient evidence that the WeRFree instrument measures capability wellbeing as it is intended to measure (Koskey, 2016).

After validating and potentially adjusting the content of the instrument, weights have to be developed to facilitate the interpretation of the WeRFree instrument’s scores. Ideally, these weights should be flexible to reflect the relative importance of each of the scales in different contexts. How such a set of weights can be developed goes beyond the current dissertation and is a question that will need to be the subject of future research.

7 CONCLUSION

The capability approach provides a promising theoretical base for the development of instruments. In the context of HTA, insights from the capability approach support the identification of relevant elements of wellbeing that are not included in conventional health economic analysis. Still, there are conceptual and methodological challenges in the application of the approach, some of which I addressed in this dissertation. Conceptually, this dissertation shows how a comprehensive concept of capability facilitates the identification of relevant elements of wellbeing that might be missed by using less comprehensive conceptualizations. Methodologically, this dissertation shows how methods that are used in psychometric research can be applied to provide insights into conceptual problems, such as the role of subjective wellbeing in self-report instruments based on the capability approach and how to assess whether adapted preferences affect instrument responses. Beyond the development of the WeRFree instrument, these findings should support the future development of the capability approach in health economics.

8 SUMMARY

Health technology assessment agencies have the task to assess the value of health technologies. In this context, some agencies use information from instruments that are completed by patients to provide evidence for the effect of health technologies. Conventional instruments that are used for this purpose have however been critiqued to be too narrow in their scope since their content is focused on measuring health-related quality of life. Proponents of the capability approach argue that health technologies have an impact on individuals' lives beyond health. They argue that health technologies improve the freedom of individuals to do and be. This freedom is also called capability. Proponents of the capability approach have developed instruments that assess wellbeing in terms of capabilities to comprehensively measure the effect of health technologies. There are however conceptual challenges in the operationalization of the capability approach into an instrument. The first study in this dissertation is a literature review of existing capability instruments to review how different authors have interpreted and operationalized the capability concept into an instrument. One conclusion of this review is that there is a large variation in how capability is measured with different instruments. Some of the content of these instruments seem to be measuring something else than capability. This results in difficulties in the interpretation of the results of these instruments. Furthermore, some instruments seem to be missing content about the burdens that people experience in their lives, such as the experience of pain. This means that these instruments might be unable to comprehensively assess the effects of health technologies. The main recommendation of this study is to use a more comprehensive and precise definition of capability to develop instruments. This would support the identification of burdens that individuals experience and facilitate a clear classification of elements of wellbeing.

Based on discussions in the literature, I propose that the concept of option freedom is such a comprehensive and precise definition of capability. This concept is operationalized into an instrument that can be used to assess capability wellbeing to illustrate the benefits of using option freedom for instrument development. The process of operationalization consists of two main stages. In the first stage, a best-fit framework synthesis was conducted to develop a theoretical framework that can be used as the basis for an instrument. In this synthesis, the concept of option freedom was applied

as an a priori concept to qualitative studies that support the development of content for capability instruments.

In the second stage, psychometric methods were used to develop an instrument that was based on this theoretical framework. This instrument consists of three scales that measure different elements of wellbeing. The first scale is called “Perceived Access to Options” and assesses the health-related capability of individuals to access options. The scales “Affective Wellbeing” and “Reflective Wellbeing” assess the subjective wellbeing derived from being able to exercise those options. Together, these three scales comprehensively assess capability wellbeing. This chapter illustrates the benefits of using the concept of option freedom for instrument development.

One further challenge with the use of self-report instruments is that responses might be affected by adapted preferences. Disadvantaged individuals might report being better off than one would expect. In the context of this discussion, I propose that measurement invariance analysis is a promising method to study if adapted preferences affect responses to instruments. The aim of measurement invariance analysis is to establish whether groups of people with different characteristics respond similarly to an instrument. Establishing the measurement invariance properties of an instrument between an advantaged and disadvantaged group (e.g., between healthy and diseased individuals) would indicate that responses are not affected by adapted preferences in the group that experiences disadvantages.

To conclude, this dissertation advances research in applying the capability approach in health economics in two ways. Firstly, it shows the importance of using comprehensive and precise concepts in capability instrument development, since these concepts guide what we as researchers include as the content in our instruments. Secondly, it shows that measurement invariance testing can be a useful tool in establishing whether instrument responses are unaffected by adapted preferences. These findings are directly relevant to patients, given the increasing use of capability instruments to assess the effect and value of health technologies.

9 REFERENCES

Al-Janabi, H., N Flynn, T., and Coast, J. (2012). Development of a self-report measure of capability wellbeing for adults: The ICECAP-A. Quality of life research *21*, 167-176.

Albrecht, G.L., and Devlieger, P.J. (1999). The disability paradox: High quality of life against all odds. Social science & medicine *48*, 977-988.

Alkire, S. (2002). Dimensions of human development. World development *30*, 181-205.

Azizli, N., Atkinson, B.E., Baughman, H.M., and Giammarco, E.A. (2015). Relationships between general self-efficacy, planning for the future, and life satisfaction. Personality and individual differences *82*, 58-60.

Bandalos, D.L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. Structural equation modeling: a multidisciplinary journal *21*, 102-116.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. Annual review of psychology *52*, 1-26.

Bandura, A. (2008). An agentic perspective on positive psychology. In Positive psychology: Exploring the best in people, S.J. Lopez, ed. (London: Praeger Publishers), pp. 167-196.

Banta, D. (2003). The development of health technology assessment. Health policy *63*, 121-132.

Banta, D. (2009). What is technology assessment? International journal of technology assessment in health care *25*, 7-9.

Bechger, T.M., Maris, G., Verstralen, H.H., and Béguin, A.A. (2003). Using classical test theory in combination with item response theory. Applied psychological measurement *27*, 319-334.

Bellanca, N., Biggeri, M., and Marchetta, F. (2011). An extension of the capability approach: Towards a theory of dis-capability. Alter *5*, 158-176.

Bentham, J. (1781). Of the principle of utility. In An Introduction to the principles of morals and legislation (Kitchener: Batoche Books, 2000), pp. 14-18.

Bentler, P.M., and Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological bulletin *88*, 588.

Berlin, I. (1969). Four essays on liberty. In Two concepts of liberty (Oxford: Oxford University Press), pp. 118-172.

Blanca, M.J., Alarcón, R., Arnau, J., Bono, R., and Bendayan, R. (2018). Effect of variance ratio on ANOVA robustness: Might 1.5 be the limit? Behavior research methods *50*, 937-962.

Boateng, G.O., Neilands, T.B., Frongillo, E.A., Melgar-Quiñonez, H.R., and Young, S.L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. Frontiers in public health *6*, 149.

Bordage, G. (2009). Conceptual frameworks to illuminate and magnify. Medical education *43*, 312-319.

Brazier, J., Roberts, J., and Deverill, M. (2002). The estimation of a preference-based measure of health from the sf-36. Journal of health economics *21*, 271-292.

Breidert, C., Hahsler, M., and Reutterer, T. (2006). A review of methods for measuring willingness-to-pay. Innovative Marketing *2*, 8-32.

Brouwer, W.B., Culyer, A.J., van Exel, N.J.A., and Rutten, F.F. (2008). Welfarism vs. Extra-welfarism. Journal of health economics *27*, 325-338.

Brouwer, W.B., and Koopmanschap, M.A. (2000). On the economic foundations of CEA. Ladies and gentlemen, take your positions! Journal of health economics *19*, 439-459.

Browne, M.W., and Cudeck, R. (1992). Alternative ways of assessing model fit. Sociological methods & research *21*, 230-258.

Busseri, M.A., and Sadava, S.W. (2011). A review of the tripartite structure of subjective well-being: Implications for conceptualization, operationalization, analysis, and synthesis. Personality and social psychology review *15*, 290-314.

Carroll, C., and Booth, A. (2015). Quality assessment of qualitative evidence for systematic review and synthesis: Is it meaningful, and if so, how should it be performed? Research synthesis methods *6*, 149-154.

Carroll, C., Booth, A., Leaviss, J., and Rick, J. (2013). “Best fit” framework synthesis: Refining the method. BMC medical research methodology *13*, 37.

Carter, I. (2003). Positive and negative liberty. In The Stanford encyclopedia of philosophy, E.N. Zalta, ed. (spring 2022 edition).

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural equation modeling: a multidisciplinary journal *14*, 464-504.

Clark, D.A. (2005). Sen's capability approach and the many spaces of human well-being. The journal of development studies *41*, 1339-1368.

Clark, L.A., and Watson, D. (1995). Constructing validity: Basic issues in objective scale development. Psychological assessment *7*, 309-319

Coast, J. (2019). Assessing capability in economic evaluation: A life course approach? The European journal of health economics *20*, 779-784.

Coast, J., Bailey, C., Orlando, R., Armour, K., Perry, R., Jones, L., and Kinghorn, P. (2018). Adaptation, acceptance and adaptive preferences in health and capability well-

being measurement amongst those approaching end of life. The patient-patient-centered outcomes research *11*, 539-546.

Coast, J., Flynn, T.N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J.J., and Peters, T.J. (2008a). Valuing the ICECAP capability index for older people. Social science & medicine *67*, 874-882.

Coast, J., Smith, R., and Lorgelly, P. (2008b). Should the capability approach be applied in health economics? Health economics *17*, 667-670.

Coast, J., Smith, R.D., and Lorgelly, P. (2008c). Welfarism, extra-welfarism and capability: The spread of ideas in health economics. Social science & medicine *67*, 1190-1198.

Comrey, A.L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. Journal of consulting and clinical psychology *56*, 754-761.

Connell, J., Brazier, J., O'Cathain, A., Lloyd-Jones, M., and Paisley, S. (2012). Quality of life of people with mental health problems: A synthesis of qualitative research. Health and quality of life outcomes *10*, 138.

Cookson, R. (2005). QALYs and the capability approach. Health economics *14*, 817-829.

Costello, A.B., and Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical assessment, research, and evaluation *10*, 1-9.

Cox, R.M., and Alexander, G.C. (1995). The abbreviated profile of hearing aid benefit. Ear and hearing *16*, 176-186.

Culyer, A.J. (1989). The normative economics of health care finance and provision. Oxford review of economic policy *5*, 34-58.

Cummins, R.A., Eckersley, R., Pallant, J., Van Vugt, J., and Misajon, R. (2003). Developing a national index of subjective wellbeing: The Australian unity wellbeing index. Social indicators research *64*, 159-190.

Cutler, D.M., and Lleras-Muney, A. (2006). Education and health: Evaluating theories and evidence (Cambridge: National Bureau of Economic Research). < https://www.nber.org/papers/w12352>

Davis, J.C., Liu-Ambrose, T., Richardson, C.G., and Bryan, S. (2013). A comparison of the ICECAP-O with EQ-5D in a falls prevention clinical setting: Are they complements or substitutes? Quality of life research *22*, 969-977.

De Castro, E.K., Ponciano, C., Meneghetti, B., and Kreling, M. (2012). Quality of life, self-efficacy and psychological well-being in Brazilian adults with cancer: A longitudinal study. Psychology *3*, 304-309.

De Champlain, A.F. (2010). A primer on classical test theory and item response theory for assessments in medical education. Medical education *44*, 109-117.

Decancq, K., and Lugo, M.A. (2013). Weights in multidimensional indices of wellbeing: An overview. Econometric reviews *32*, 7-34.

DeVellis, R.F. (2017a). Factor analysis. In Scale development: Theory and applications, 4th edition, L. Bickman and D.J. Rog, eds. (Los Angeles: SAGE Publications), pp. 142-191.

DeVellis, R.F. (2017b). Guidelines in scale development. In Scale development: Theory and applications, 4th edition, L. Bickman and D.J. Rog, eds. (Los Angeles: SAGE Publications), pp. 121-123.

DeVellis, R.F. (2017c). Reliability. In Scale development: Theory and applications, 4th edition, L. Bickman and D.J. Rog, eds. (Los Angeles: SAGE Publications), pp. 49-85.

DeVellis, R.F. (2017d). Understanding the latent variable. In Scale development: Theory and applications, 4th edition, L. Bickman and D.J. Rog, eds. (Los Angeles: SAGE Publications), pp. 36-48.

DeVellis, R.F. (2017e). Validity. In Scale development: Theory and applications, 4th edition, L. Bickman and D.J. Rog, eds. (Los Angeles: SAGE Publications), pp. 86 - 102.

Diener, E. (1984). Subjective well-being. Psychological Bulletin *95*, 542-575.

Diener, E., Emmons, R.A., Larsen, R.J., and Griffin, S. (1985). The satisfaction with life scale. Journal of personality assessment *49*, 71-75.

Diener, E., Scollon, C.N., and Lucas, R.E. (2009). The evolving concept of subjective well-being: The multifaceted nature of happiness. In Assessing well-being, E. Diener, ed. (Dordrecht: Springer) pp. 67-100

Dolan, P., and Metcalfe, R. (2012). Measuring subjective wellbeing: Recommendations on measures for use by national governments. Journal of social policy *41*, 409-427.

Dolan, P., Peasgood, T., and White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. Journal of economic psychology *29*, 94-122.

Dolan, P., and White, M.P. (2007). How can measures of subjective well-being be used to inform public policy? Perspectives on psychological science *2*, 71-85.

Driscoll, W.C. (1996). Robustness of the ANOVA and Tukey-Kramer statistical tests. Computers & industrial engineering *31*, 265-268.

Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L., and Torrance, G.W. (2015a). Introduction to economic evaluation. In Methods for the economic evaluation of health care programmes (Oxford: Oxford university press), pp. 1-18.

Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L., and Torrance, G.W. (2015b). Measuring and valuing effects: Health gain. In Methods for the economic evaluation of health care programmes (Oxford university press), pp. 123-180.

Eckford, R.D., Gaisser, A., Arndt, V., Baumann, M., Kludt, E., Mehlis, K., Ubels, J., Winkler, E.C., Weg-Remers, S., and Schlander, M. (2021). The COVID-19 pandemic and cancer patients in Germany: Impact on treatment, follow-up care and psychological burden. Frontiers in public health *9*.

Enders, C.K., and Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Structural equation modeling *8*, 430-457.

Engel, L., Mortimer, D., Bryan, S., Lear, S.A., and Whitehurst, D.G. (2017). An investigation of the overlap between the ICECAP-A and five preference-based health-related quality of life instruments. PharmacoEconomics *35*, 741-753.

Engström, M.S., Leksell, J., Johansson, U.B., Eeg-Olofsson, K., Borg, S., Palaszewski, B., and Gudbjörnsdottir, S. (2018). A disease-specific questionnaire for measuring patient-reported outcomes and experiences in the Swedish national diabetes register: Development and evaluation of content validity, face validity, and test-retest reliability. Patient education and counseling *101*, 139-146.

Engström, M.S., Leksell, J., Johansson, U.B., and Gudbjörnsdottir, S. (2016). What is important for you? A qualitative interview study of living with diabetes and experiences of diabetes care to establish a basis for a tailored patient-reported outcome measure for the Swedish national diabetes register. BMJ Open *6*.

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. Educational and psychological measurement *58*, 357-381.

Fleurbaey, M. (2006). Capabilities, functionings and refined functionings. Journal of human development *7*, 299-310.

Franklin, M., Payne, K., and Elliott, R.A. (2018). Quantifying the relationship between capability and health in older people: Can't map, won't map. Medical decision making *38*, 79-94.

Frei, A., Svarin, A., Steurer-Stey, C., and Puhan, M.A. (2009). Self-efficacy instruments for patients with chronic diseases suffer from methodological limitations-a systematic review. Health and quality of life outcomes *7*, 1-10.

Gaisser, A., Eckford, R.D., Arndt, V., Doege, D., Kludt, E., Ubels, J., Schlander, M., and Weg-Remers, S. (2022). Fast zwei jahre coronapandemie aus der perspektive von krebsbetroffenen. Forum *37*, 216-220.

Gardner, P.L. (1996). The dimensionality of attitude scales: A widely misunderstood idea. International journal of science education *18*, 913-919.

Gasper, D. (2010). Understanding the diversity of conceptions of well-being and quality of life. The journal of socio-economics *39*, 351-360.

Gaynes, B.N., Burns, B.J., Tweed, D.L., and Erickson, P. (2002). Depression and health-related quality of life. The journal of nervous and mental disease *190*, 799-806.

Gibbins, J., Bhatia, R., Forbes, K., and Reid, C.M. (2014). What do patients with advanced incurable cancer want from the management of their pain? A qualitative study. Palliative medicine *28*, 71-78.

Greco, G. (2013). Assessing women's quality of life in rural Malawi: A capabilities index. (PhD Thesis, London School of Hygiene & Tropical Medicine).

Greco, G., Skordis-Worrall, J., Mkandawire, B., and Mills, A. (2015). What is a good life? Selecting capabilities to assess women's quality of life in rural Malawi. Social science & medicine *130*, 69-78.

Grewal, I., Lewis, J., Flynn, T., Brown, J., Bond, J., and Coast, J. (2006). Developing attributes for a generic quality of life measure for older people: Preferences or capabilities? Social science & medicine *62*, 1891-1901.

Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. Journal of health economics *19*, 403-420.

Hackert, M.Q., van Exel, J., and Brouwer, W.B. (2017). Valid outcome measures in care for older people: Comparing the ASCOT and the ICECAP-O. Value in health *20*, 936-944.

Hambleton, R.K., and Jones, R.W. (1993). Comparison of classical test theory and item response theory and their applications to test development. Educational measurement: Issues and practice *12*, 38-47.

Hawthorne, G., Richardson, J., and Osborne, R. (1999). The assessment of quality of life (AQoL) instrument: A psychometric measure of health-related quality of life. Quality of life research *8*, 209-224.

Helter, T.M., Coast, J., Łaszewska, A., Stamm, T., and Simon, J. (2020). Capability instruments in economic evaluations of health-related interventions: A comparative review of the literature. Quality of life research *29*, 1433-1464.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G., and Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of life research *20*, 1727-1736.

Hofmann, B. (2005). On value-judgements and ethics in health technology assessment. Poiesis & praxis *3*, 277-295.

Hörder, H., Gustafsson, S., Rydberg, T., Skoog, I., and Waern, M. (2016). A cross-cultural adaptation of the ICECAP-O: Test–retest reliability and item relevance in Swedish 70-year-olds. Societies *6*, 30.

Hu, L.t., and Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal *6*, 1-55.

Huang, I.C., Leite, W.L., Shearer, P., Seid, M., Revicki, D.A., and Shenkman, E.A. (2011). Differential item functioning in quality of life measure between children with and without special health-care needs. Value in health *14*, 872-883.

Hurley, J. (2000). An overview of the normative economics of the health sector. In Handbook of health economics, A.J. Culyer and J.P. Newhouse, eds. (Amsterdam: Elsevier), pp. 55-118.

Huta, V., and Waterman, A.S. (2014). Eudaimonia and its distinction from hedonia: Developing a classification and terminology for understanding conceptual and operational definitions. Journal of happiness studies *15*, 1425-1456.

Jang, S., Kim, E.S., Cao, C., Allen, T.D., Cooper, C.L., Lapierre, L.M., O'Driscoll, M.P., Sanchez, J.I., Spector, P.E., and Poelmans, S.A. (2017). Measurement invariance of the satisfaction with life scale across 26 countries. Journal of cross-cultural psychology *48*, 560-576.

Jeong, S., and Lee, Y. (2019). Consequences of not conducting measurement invariance tests in cross-cultural studies: A review of current research practices and recommendations. Advances in developing human resources *21*, 466-483.

Jung, E., and Yoon, M. (2016). Comparisons of three empirical methods for partial factorial invariance: Forward, backward, and factor-ratio tests. Structural equation modeling: A multidisciplinary journal *23*, 567-584.

Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica *47*, 363-391.

Kaplan, R.M., Anderson, J.P., and Ganiats, T.G. (1993). The quality of well-being scale: Rationale for a single quality of life index. In Quality of life assessment: Key issues in the 1990s, S.R. Walker, and R.M. Rosser, eds. (Springer), pp. 65-94.

Keeley, T., Coast, J., Nicholls, E., Foster, N.E., Jowett, S., and Al-Janabi, H. (2016). An analysis of the complementarity of ICECAP-A and EQ-5D-3L in an adult population of patients with knee pain. Health and quality of life outcomes *14*, 1-5.

Khader, S.J. (2011). Adaptive preferences and women's empowerment (New York: Oxford University Press).

Khader, S.J. (2013). Identifying adaptive preferences in practice: Lessons from postcolonial feminisms. Journal of global ethics *9*, 311-327.

Khan, M.A., and Richardson, J. (2018). Variation in the apparent importance of health-related problems with the instrument used to measure patient welfare. Quality of life research *27*, 2885-2896.

Kibel, M., and Vanstone, M. (2017). Reconciling ethical and economic conceptions of value in health policy using the capabilities approach: A qualitative investigation of non-invasive prenatal testing. Social science & medicine *195*, 97-104.

Kim, T.K. (2017). Understanding one-way ANOVA using conceptual figures. Korean journal of anesthesiology *70*, 22-26.

Kinghorn, P. (2015). Exploring different interpretations of the capability approach in a health care context: Where next? Journal of human development and capabilities *16*, 600-616.

Kinghorn, P., Robinson, A., and Smith, R.D. (2015). Developing a capability-based questionnaire for assessing well-being in patients with chronic pain. Social indicators research *120*, 897-916. https://doi.org/10.1007/s11205-014-0625-7

Kline, R.B. (2011a). Data preparation. In Principles and practice of structural equation modeling, 3rd edition, (New York: The Guilford Press), pp. 46 - 74.

Kline, R.B. (2011b). Estimation. In Principles and practice of structural equation modeling, 3rd edition, (New York: The Guilford Press), pp. 154-185.

Kline, R.B. (2011c). Specification. In Principles and practice of structural equation modeling, 3rd edition, (New York: The Guilford Press), pp. 91-123.

Knott, R.J., Black, N., Hollingsworth, B., and Lorgelly, P.K. (2017a). Response-scale heterogeneity in the EQ-5D. Health economics *26*, 387-394.

Knott, R.J., Lorgelly, P.K., Black, N., and Hollingsworth, B. (2017b). Differential item functioning in quality of life measurement: An analysis using anchoring vignettes. Social science & medicine *190*, 247-255.

Korkmaz, S., Göksülük, D., and Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. The R journal *6,* 151-162.

Koskey, K.L. (2016). Using the cognitive pretesting method to gain insight into participants' experiences: An illustration and methodological reflection. International journal of qualitative methods *15*, 1609406915624577.

Kramer, C.Y. (1956). Extension of multiple range tests to group means with unequal numbers of replications. Biometrics *12*, 307-310.

Kymlicka, W. (2002). Utilitarianism. In Contemporary political philosophy: An introduction, 2nd edition (Oxford: Oxford University Press), pp. 13 - 20.

Kyriazos, T.A., and Stalikas, A. (2018). Applied psychometrics: The steps of scale development and standardization process. Psychology *9*, 2531-2560.

Le, H., Schmidt, F.L., Harter, J.K., and Lauver, K.J. (2010). The problem of empirical redundancy of constructs in organizational research: An empirical investigation. Organizational behavior and human decision processes *112*, 112-125.

Lent, R.W., Singley, D., Sheu, H.-B., Gainor, K.A., Brenner, B.R., Treistman, D., and Ades, L. (2005). Social cognitive predictors of domain and life satisfaction: Exploring the theoretical precursors of subjective well-being. Journal of counseling psychology *52*, 429-442.

Lenzner, T., Neuert, C., and Otto, W. (2016). Cognitive pretesting (version 2.0). In GESIS Survey Guidelines (Mannheim: GESIS - Leibniz Institute for the Social Sciences). <https://doi.org/10.15465/gesis-sg_en_010>

Li, C.-H. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. Behavior research methods *48*, 936-949.

Li, C.-H. (2016b). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. Psychological methods *21*, 369-387.

Lincoln, Y.S., and Guba, E.G. (1985). Establishing trustworthiness. In Naturalistic inquiry (Newbury Park: SAGE Publications), pp. 289 - 331.

Linton, M.-J., Mitchell, P.M., Al-Janabi, H., Schlander, M., Richardson, J., Iezzi, A., Ubels, J., and Coast, J. (2020). Comparing the German translation of the ICECAP-A capability wellbeing measure to the original English version: Psychometric properties across healthy samples and seven health condition groups. Applied research in quality of life *15*, 651-673.

López Ulloa, B.F., Møller, V., and Sousa-Poza, A. (2013). How does subjective well-being evolve with age? A literature review. Journal of population ageing *6*, 227-246.

Lorgelly, P., Lorimer, K., Fenwick, E., and Briggs, A. (2008). The capability approach: Developing and instrument for evaluating public health interventions (Glasgow: University of Glasgow). < https://lx.iriss.org.uk/sites/default/files/resources/Capabilities%20-%20full%20report,%20August%2008.pdf>

Lorgelly, P.K., Lawson, K.D., Fenwick, E.A., and Briggs, A.H. (2010). Outcome measurement in economic evaluations of public health interventions: A role for the capability approach? International journal of environmental research and public health *7*, 2274-2289.

Lorgelly, P.K., Lorimer, K., Fenwick, E.A.L., Briggs, A.H., and Anand, P. (2015). Operationalising the capability approach as an outcome measure in public health: The development of the OCAP-18. Social science & medicine *142*, 68-81.

Lorimer, K., Lorgelly, P., Fenwick, E., and Briggs, A. (2007). Using mixed methods to operationalise the capability approach: An application in public health. In Human development and capability association conference (2007) (New York: New School University)

Lucas, R.E., Diener, E., and Suh, E. (1996). Discriminant validity of well-being measures. Journal of personality and social psychology *71*, 616-628.

Luhmann, M., Hofmann, W., Eid, M., and Lucas, R.E. (2012). Subjective well-being and adaptation to life events: A meta-analysis. Journal of personality and social psychology *102*, 592-615.

Macdonald, P., and Paunonen, S.V. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. Educational and psychological measurement *62*, 921-943.

MacKenzie, S.B., and Podsakoff, P.M. (2012). Common method bias in marketing: Causes, mechanisms, and procedural remedies. Journal of retailing *88*, 542-555.

Månsdotter, A., Ekman, B., Feldman, I., Hagberg, L., Hurtig, A.-K., and Lindholm, L. (2017). We propose a novel measure for social welfare and public health: Capability-adjusted life-years, CALYs. Applied health economics and health policy *15*, 437-440.

Månsdotter, A., Ekman, B., Meili, K.W., Feldman, I., Hagberg, L., Hurtig, A.-K., and Lindholm, L. (2020). Towards capability-adjusted life years in public health and social welfare: Results from a Swedish survey on ranking capabilities. PloS one *15*, e0242699.

Marks, R., and Allegrante, J.P. (2005). A review and synthesis of research evidence for self-efficacy-enhancing interventions for reducing chronic disability: Implications for health education practice (part II). Health promotion practice *6*, 148-156.

Martinez-Calderon, J., Meeus, M., Struyf, F., and Luque-Suarez, A. (2020). The role of self-efficacy in pain intensity, function, psychological factors, health behaviors, and quality of life in people with rheumatoid arthritis: A systematic review. Physiotherapy theory and practice *36*, 21-37.

Meredith, W., and Teresi, J.A. (2006). An essay on measurement and factorial invariance. Medical care *44*, 69-77.

Mill, J.S. (1887). Utilitarianism (Boston: Willard Small).

Millsap, R.E. (2007). Invariance in measurement and prediction revisited. Psychometrika *72*, 461-473.

Mitchell, P. (2018). Adaptive preferences, adapted preferences. Mind *127*, 1003-1025.

Mitchell, P.M., Roberts, T.E., Barton, P.M., and Coast, J. (2015). Assessing sufficient capability: A new approach to economic evaluation. Social science & medicine *139*, 71-79.

Mitchell, P.M., Roberts, T.E., Barton, P.M., and Coast, J. (2017). Applications of the capability approach in the health field: A literature review. Social indicators research *133*, 345-371.

Monitorering & Medicinsk Teknologivurdering (2009). Kontrolforløb for gynækologiske kræftpatienter - en medicinsk teknologivurdering (Copenhagen: Sundhedsstyrelsen, Monitorering & Medicinsk Teknologivurdering). < https://www.sst.dk/Udgivelser/2009/~/media/EB99C23F419F4CF2B198387B7AFAE5A2.ashx>

Moore, A. (2004). Hedonism. In The Stanford encyclopedia of philosophy (Winter 2019 Edition), E.N. Zalta ed. (Stanford: Stanford University).

Netten, A., Burge, P., Malley, J., Potoglou, D., Towers, A.-M., Brazier, J., Flynn, T., and Forder, J. (2012). Outcomes of social care for adults: Developing a preference-weighted measure. Health technology assessment *16*, 1-166.

Nussbaum, M. (2003). Capabilities as fundamental entitlements: Sen and social justice. Feminist Economics *9*, 35-59.

Sen, A. (1993a). Capability and wellbeing. In The quality of life, M. Nussbaum and A. Sen, eds. (Oxford: Oxford University Press), p. 31.

Sen, A. (1993b). Capability and wellbeing. In The quality of life, M. Nussbaum and A. Sen, eds. (Oxford: Oxford University Press), p. 39.

Pettit, P. (2003). Agency-freedom and option-freedom. Journal of theoretical politics *15*, 387-403.

Phillips, D. (2006). Quality of life: Concept, policy and practice (London: Routledge).

Progar, Š., Sočan, G., and Peč, M. (2008). An empirical comparison of item response theory and classical test theory. Horizons of psychology *17*, 5-24.

Putnick, D.L., and Bornstein, M.H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Developmental review *41*, 71-90.

Qizilbash, M. (2002). Development, common foes and shared values. Review of Political Economy *14*, 463-480.

R Core Team (2023). R: A language and environment for statistical computing. < https://www.R-project.org/>

Ramsey, P.H., Ramsey, P.P., and Barrera, K. (2010). Choosing the best pairwise comparisons of means from non-normal populations, with unequal variances, but equal sample sizes. Journal of statistical computation and simulation *80*, 595-608.

Reise, S.P. (2012). The rediscovery of bifactor measurement models. Multivariate behavioral research *47*, 667-696.

Rencz, F., Mitev, A.Z., Jenei, B., and Brodszky, V. (2021). Measurement properties of the ICECAP-A capability well-being instrument among dermatological patients. Quality of life research *31*, 903-915.

Revelle, W. (2023). Psych: Procedures for psychological, psychometric, and personality research. Northwestern University, Evanston, Illinois. <https://CRAN.R-project.org/package=psych>.

Rhemtulla, M., Brosseau-Liard, P.É., and Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychological methods *17*, 354-373.

Richardson, J., and Iezzi, A. (2014). Do utility formulate accentuate or diminish differences between multi attribute utility (MAUI) instruments (Melbourne: Monash University). <https://aqol.com.au/papers/workingpaper03-2014.pdf>

Richardson, J., Iezzi, A., and Khan, M.A. (2015). Why do multi-attribute utility instruments produce different utilities: The relative importance of the descriptive systems, scale and 'micro-utility' effects. Quality of life research *24*, 2045-2053.

Richardson, J., Iezzi, A., Khan, M.A., Chen, G., and Maxwell, A. (2016). Measuring the sensitivity and construct validity of 6 utility instruments in 7 disease areas. Medical decision making *36*, 147-159.

Richardson, J., Khan, M., Iezzi, A., and Maxwell, A. (2012). Cross-national comparison of twelve quality of life instruments: MIC paper 1: background, questions, instruments. (Melbourne: Monash University). <https://www.aqol.com.au/papers/researchpaper76.pdf>

Richardson, J., Khan, M., Iezzi, A., Sinha, K., Mihalopoulos, C., Herrman, H., Hawthorne, G., and Schweitzer, I. (2009). The AQoL-8D (PsyQol) MAU instrument: Overview September 2009. (Melbourne: Monash University).

Rijke, W.J., Vermeulen, A.M., Wendrich, K., Mylanus, E., Langereis, M.C., and van der Wilt, G.J. (2019). Capability of deaf children with a cochlear implant. Disability and rehabilitation *43*, 1989-1994

Robeyns, I. (2003). Sen's capability approach and gender inequality: Selecting relevant capabilities. Feminist economics *9*, 61-92.

Robeyns, I. (2005). Selecting capabilities for quality of life measurement. Social indicators research *74*, 191-215.

Robeyns, I. (2006). The capability approach in practice. Journal of political philosophy *14*, 351-376.

Robeyns, I. (2017a). Clarifications: Capabilities as opportunity or option freedoms?. In Wellbeing, freedom and social justice: The capability approach re-examined (Cambridge: Open Book Publishers), pp. 102 - 105.

Robeyns, I. (2017b). Clarifications. In Wellbeing, freedom and social justice: The capability approach re-examined (Cambridge Open Book Publishers), pp. 137-142.

Robeyns, I. (2017c). Clarifications: Are capabilities freedoms, and if so, which ones? In Wellbeing, freedom and social justice: The capability approach re-examined (Cambridge: Open Book Publishers), pp. 98 - 107.

Robeyns, I. (2017d). Core ideas and the framework: The C-modules: contingent modules. In Wellbeing, freedom and social justice: The capability approach re-examined (Cambridge: Open Book Publishers), pp. 67 - 74.

Rodriguez, A., Reise, S.P., and Haviland, M.G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. Psychological methods *21*, 137.

Rohrbach, P.J., Dingemans, A.E., Essers, B.A., Van Furth, E.F., Spinhoven, P., Groothuis-Oudshoorn, C.G., Van Til, J.A., den Akker-Van Marle, V., and Elske, M. (2021). The ICECAP-A instrument for capabilities: Assessment of construct validity and test–retest reliability in a general Dutch population. Quality of Life Research *31*, 687–696.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Journal of statistical software *48*, 1-36.

Rosseel, Y. (2014). The Lavaan tutorial. https://lavaan.ugent.be/tutorial/. Accessed January 2023.

Rudmik, L., and Drummond, M. (2013). Health economic evaluation: Important principles and methodology. The laryngoscope *123*, 1341-1347.

Ryan, R.M., and Deci, E.L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. Annual review of psychology *52*, 141-166.

Ryff, C.D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. Journal of personality and social psychology *57*, 1069-1081.

Sajobi, T.T., Brahmbatt, R., Lix, L.M., Zumbo, B.D., and Sawatzky, R. (2018). Scoping review of response shift methods: Current reporting practices and recommendations. Quality of life research *27*, 1133-1146.

Sanghera, S., Walther, A., Peters, T.J., and Coast, J. (2022). Challenges in using recommended quality of life measures to assess fluctuating health: A think-aloud study to understand how recall and timing of assessment influence patient responses. The patient-Patient-centered outcomes research *15*, pages 445–457.

Scanlon, T.M. (1991). The moral basis of interpersonal comparisons. In Interpersonal comparisons of wellbeing, J. Elster and J.E. Roemer, eds. (Cambridge: Cambridge University Press), pp. 17-44.

Schaefer, R., Hernandez, D., Selberg, L., and Schlander, M. (2021). Health technology assessment (HTA) in England, France and Germany: What do matched drug pairs tell us about recommendations by national HTA agencies? Journal of comparative effectiveness research *10*, 1187-1195.

Schaefer, R., and Schlander, M. (2019). Is the national institute for health and care excellence (NICE) in England more 'innovation-friendly' than the federal joint committee (G-BA) in Germany? Expert review of pharmacoeconomics & outcomes research *19*, 453-462.

Schlander, M. (2005). Kosteneffektivität und Ressourcenallokation: Gibt es einen normativen Anspruch der Gesundheitsökonomie. In Gesundheitswesen zwischen Wirtschaftlichkeit und Menschlichkeit, H.A. Kick and J. Taupitz, eds. (Münster: LIT-Verlag), pp. 37-112.

Schlosser, R.W., Wendt, O., Bhavnani, S., and Nail-Chiwetalu, B. (2006). Use of information-seeking strategies for developing systematic reviews and engaging in

evidence-based practice: The application of traditional and comprehensive pearl growing. A review. International journal of language & communication disorders *41*, 567-582.

Schönfeld, P., Brailovskaia, J., Bieda, A., Zhang, X.C., and Margraf, J. (2016). The effects of daily stress on positive and negative mental health: Mediation through self-efficacy. International journal of clinical and health psychology *16*, 1-10.

Schroeder, M. (2021). Value theory. In The Stanford encyclopedia of philosophy (Fall 2021 edition), E.N. Zalta, ed. (Stanford: Stanford University).

Schwartz, C.E., Bode, R., Repucci, N., Becker, J., Sprangers, M.A., and Fayers, P.M. (2006). The clinical significance of adaptation to changing health: A meta-analysis of response shift. Quality of life research *15*, 1533-1550.

Seiber, W.J., Groessl, E.J., David, K.M., Ganiats, T.G., and Kaplan, R.M. (2008). Quality of well being self-administered (QWB-SA) scale (San Diego: University of California). < https://hoap.ucsd.edu/qwb-info/QWB-Manual.pdf>

Sen, A. (1984). The living standard. Oxford economic papers *36*, 74-90.

Sen, A. (1985a). Well-being, agency and freedom: The Dewey lectures 1984. The journal of philosophy *82*, 169-221.

Sen, A. (1985b). Well-being, agency and freedom: The Dewey lectures 1984. The journal of philosophy *82*, 204-208.

Sen, A. (1987). The standard of living. G. Hawthorn, ed. (Cambridge: Cambridge University Press).

Sen, A. (1992a). Freedom, achievement and resources. In Inequality reexamined (Oxford: Oxford University Press), p. 31.

Sen, A. (1992b). Functionings and capabilities. In Inequality reexamined (Oxford: Oxford University Press), pp. 53-55.

Sen, A. (1992c). Functionings and capabilities. In Inequality reexamined (Oxford: Oxford University Press), pp. 41-42.

Sen, A. (1992d). Functionings and capabilities. In Inequality reexamined (Oxford: Oxford University Press), pp. 49-55.

Sen, A. (1997). Introduction. In Choice, welfare and measurement (Cambridge: Harvard University Press), p. 30.

Sen, A. (1999a). Commodities and capabilities (New Delhi: Oxford University Press)

Sen, A. (1999b). Commodities and capabilities (New Delhi: Oxford University Press), 1 - 5.

Sen, A. (2002). Health: Perception versus observation: Self reported morbidity has severe limitations and can be extremely misleading. BMJ *324,* 860-861.

Sen, A. (2004). Capabilities, lists, and public reason: Continuing the conversation. Feminist economics *10*, 77-80.

Simms, L.J., Zelazny, K., Williams, T.F., and Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. Psychological assessment *31*, 557-566.

Simon, J., Anand, P., Gray, A., Rugkåsa, J., Yeeles, K., and Burns, T. (2013). Operationalising the capability approach for outcome measurement in mental health research. Social science & medicine *98*, 187-196.

Sintonen, H., and Pekurinen, M. (1993). A fifteen-dimensional measure of health-related quality of life (15D) and its applications. In Quality of life assessment: Key issues in the 1990s, S.R. Walker and R.M. Rosser eds. (Dordrecht: Springer), pp. 185-195.

Smith, A.B., Cocks, K., Parry, D., and Taylor, M. (2016). A differential item functioning analysis of the eq-5d in cancer. Value in health *19*, 1063-1067.

Smith, D.M., Sherriff, R.L., Damschroder, L., Loewenstein, G., and Ubel, P.A. (2006). Misremembering colostomies? Former patients give lower utility ratings than do current patients. Health psychology *25*, 688-695.

Sprangers, M.A., and Schwartz, C.E. (1999). Integrating response shift into health-related quality of life research: A theoretical model. Social science & medicine *48*, 1507-1515.

Strobel, M., Tumasjan, A., and Spörrle, M. (2011). Be yourself, believe in yourself, and be happy: Self-efficacy as a mediator between personality factors and subjective well-being. Scandinavian journal of psychology *52*, 43-48.

Sutton, E.J., and Coast, J. (2014). Development of a supportive care measure for economic evaluation of end-of-life care using qualitative methods. Palliative medicine *28*, 151-157.

Till, M., Abu-Omar, K., Ferschl, S., Reimers, A.K., and Gelius, P. (2021). Measuring capabilities in health and physical activity promotion: A systematic review. BMC Public Health *21*, 1-23.

Tinsley, H.E., and Tinsley, D.J. (1987). Uses of factor analysis in counseling psychology research. Journal of counseling psychology *34*, 414-424.

Tong, A., Flemming, K., McInnes, E., Oliver, S., and Craig, J. (2012). Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC medical research methodology *12*, 181-188.

Tong, A., Sainsbury, P., and Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. International journal for quality in health care *19*, 349-357.

Torrance, G.W. (1986). Measurement of health state utilities for economic appraisal: A review. Journal of health economics *5*, 1-30.

Ubel, P.A., Loewenstein, G., Schwarz, N., and Smith, D. (2005). Misimagining the unimaginable: The disability paradox and health care decision making. Health psychology *24*, 57-62.

Ubels, J., Hernandez-Villafuerte, K., Niebauer, E., and Schlander, M. (2022a). The value of freedom: Extending the evaluative space of capability. MedRxiv, 2022.2006. 2029.22277019.

Ubels, J., Hernandez-Villafuerte, K., and Schlander, M. (2022b). The value of freedom: A review of the current developments and conceptual issues in the measurement of capability. Journal of human development and capabilities *23*, 327-353.

Ubels, J., Hernandez-Villafuerte, K., and Schlander, M. (2022c). The value of freedom: The development of the WeRFree capability instrument. MedRxiv, 2022.2010. 2005.22280720.

Vallerand, A.H., Saunders, M.M., and Anthony, M. (2007). Perceptions of control over pain by patients with cancer and their caregivers. Pain management nursing *8*, 55-63.

van Leeuwen, K.M., Jansen, A.P., Muntinga, M.E., Bosmans, J.E., Westerman, M.J., van Tulder, M.W., and van der Horst, H.E. (2015). Exploration of the content validity and feasibility of the EQ-5D-3L, ICECAP-O and ASCOT in older adults. BMC health services research *15*, 1-10.

Versteegh, M., and Brouwer, W. (2016). Patient and general public preferences for health states: A call to reconsider current guidelines. Social science & medicine *165*, 66-74.

Victorson, D., Barocas, J., Song, J., and Cella, D. (2008). Reliability across studies from the functional assessment of cancer therapy-general (FACT-G) and its subscales: A reliability generalization. Quality of life research *17*, 1137-1146.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H.L., and Kievit, R.A. (2012). An agenda for purely confirmatory research. Perspectives on psychological science *7*, 632-638.

Widaman, K.F., and Reise, S.P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In The science of prevention: Methodological advances from alcohol and substance abuse research, K.J. Bryant, M. Windle, and S.G. West, eds. (Washington: American Psychological Association).

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., and Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. Value in health *8*, 94-104.

Williams, A. (1997). Intergenerational equity: An exploration of the 'fair innings' argument. Health economics *6*, 117-132.

Zeppetella, G. (1999). How do terminally ill patients at home take their medication? Palliative medicine *13*, 469-475.

10 APPENDIX

10.1 Appendix to Chapter 2. Tables and figures.

Appendix Table 1. Detailed descriptive information of identified capability instruments

Instrument	Authors	Development stage	Location	Target population	Number of informants to develop content	Method
ASCOT	Netten et al. (2012)	Instrument is developed and validated	England	Population in contact with social care services	30	Further refinement of an existing instrument with experts and cognitive interviews with participants in contact with social care services, in order to select and refine domains and questions
Capability Adjusted Life Year (CALY)	Månsdotter et al. (Månsdotter et al., 2017; Månsdotter et al., 2020)	In development	Sweden	Swedish general population	In case of Delphi study unclear. For ranking 167 respondents.	Refine and rank a list of pre-defined domains using quantitative methods and a Delphi study.
Capability-Based Questionnaire for Assessing Well-Being in Patients with Chronic Pain	Kinghorn et al. (2015)	Domains and items identified	England	People affected by chronic pain	22	In the first stage 6 focus group discussions and one interview, in the second stage results were presented back to sub-sample of initial participants with interviews
Child – and parent report questionnaire to explore capability of deaf children wearing a cochlear implant	Rijke et al. (2019)	Domains and items identified	Netherlands	Deaf children	Unclear	Content derived from literature, conversations with parents of children with cochlear implants, and experts who work at a cochlear implant team.
Diabetes specific instrument for measuring patient reported outcomes and experiences in the Swedish National Diabetes Register	Engström et al. (Engström et al., 2018; Engström et al., 2016)	Instrument developed, is being further validated	Sweden	Adults affected by diabetes	29	Interviews with an interview guide
ICECAP-A	Al-Janabi et al. (2012)	Instrument is developed and validated	England	General population	36	Interviews with topic guide, results were presented back to

Instrument	Authors	Development stage	Location	Target population	Number of informants to develop content	Method
						sub-sample of initial participants with interviews
ICECAP-SCM	Sutton and Coast (2014)	Instrument developed, is being further validated	England	65 + receiving supportive care	24	Interviews, with topic guide, results were presented back to sub-sample of initial participants with interviews
ICECAP-O	Grewal et al. (2006)	Instrument is developed and validated	England	Older people, 65+	40	In-depth interviews with broad questioning, results were then presented back to sub-sample of initial participants with interviews
Non-invasive prenatal testing elated capability wellbeing	Kibel and Vanstone (2017)	Domains identified	Canada	Adult Women	38	Secondary analysis of interviews, using Nussbaum's list of capabilities
OCAP-18	Lorgelly et al. (Lorgelly et al., 2008; Lorgelly et al., 2015)	Instrument is developed	Scotland	General population	Approx. 40 for focus groups, 17 for interviews	Cognitive interviews, 5 focus group discussions, and quantitative methods to select and refine questions from a larger questionnaire, which aim to measure the capabilities from Nussbaum's list
OxCap-MH	Simon et al. (2013)	Instrument is developed	England	People affected by mental health problems	20	Further refinement of the OCAP-18 with expert focus group discussions, interviews of participants in trial and quantitative methods, in order to select and adapt questions for use in mental health patients
Women's capability index	Greco et al. (Greco, 2013; Greco et al., 2015)	Instrument developed, is being further validated	Malawi	Adult Women	129	16 focus group discussions with topic guide

Table has been published previously in Ubels et al. (2022b).

Appendix Table 2. Content of the instrument compared to elements of option freedom, functioning and other content which could not be grouped

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
ASCOT	Could you tell me which of the following statements best describes how much control you have over your daily life? (4 point scale) • I have as much control over my daily life as I want • I have no control over my daily life		X		X	
	Do the support and services that you get from Social Services help you to maintain control over your daily life? • Yes • No • Don't know	X	X			
	Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. In that situation, which of the following would best describe the amount of control you'd have over your daily life? (4 point scale) • I would have as much control over my daily life as I want • I would have no control over my daily life	X	X		X	
	Thinking about your personal care, by which we mean being clean and presentable in		X		X	

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	appearance, which of the following statements best describes your situation? (4 point scale) • I feel clean and am able to present myself the way I like • I don't feel at all clean or presentable					
	Do the support and services that you get from Social Services help you to stay clean and presentable? • Yes • No • Don't know	X	X			
	Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. Which of the following would then best describe your situation with regard to your personal care? (4 point scale) • I would feel clean and would be able to present myself the way I like • I wouldn't feel at all clean or presentable	X	X		X	
	Thinking about the food and drink you have, which of the following statements best describes your situation? (4 point scale) • I get all the food and drink I like when I want • I don't get all the food and drink I need, and I think there is a risk to my health	X	X			X
	Do the support and services that you get from Social Services help you to get the food and drink you want or need? • Yes • No	X	X			

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• Don't know					
	Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. Which of the following would then best describe your situation with regard to food and drink? (4 point scale) • I would get all the food and drink I like when I want • I wouldn't get all the food and drink I need, and I think there would be a risk to my health	X	X		X	
	Could you tell me which of the following statements best describes how clean and comfortable your home is? (4 point scale) • My home is as clean and comfortable as I want • My home is not at all clean or comfortable		X		X	
	Do the support and services that you get from Social Services help you to keep your home clean and comfortable? • Yes • No • Don't know •	X	X			
	Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. In that situation, which of the following would best describe how clean and comfortable your home is? (4 point scale) • My home would be as clean and comfortable as I want	X	X		X	

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• My home would not be at all clean or comfortable					
	Could you tell me which of the following statements best describes how safe you feel? (4 point scale) • I feel as safe as I want • I don't feel at all safe				X	
	Do the support and services that you get from Social Services help you to feel safe? • Yes • No • Don't know	X	X			
	Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. In that situation, which of the following would best describe how safe you feel? (4 point scale) • I would feel as safe as I want • I wouldn't feel at all safe	X	X			
	Thinking about how much contact you've had with people you like, which of the following statements best describes your social situation? (4 point scale) • I have as much social contact as I want with people I like • I have little social contact with people and feel socially isolated		X		X	

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	Do the support and services that you get from Social Services help you to have contact with people you like? • Yes • No • Don't know	X	X			
	Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. In that situation, which of the following would best describe how much contact you have with people you like? (4 point scale) • I would have as much social contact as I want with people I like • I would have little social contact with people and would feel socially isolated	X	X			
	Could you tell me which of the following statements best describes how you spend your time? (4 point scale) • I'm able to spend my time as I want, doing things I value or enjoy • I don't do anything I value or enjoy with my time		X		X	
	Do the support and services that you get from Social Services help you to spend your time doing things you value and enjoy? • Yes • No • Don't know	X	X			
	Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. In that situation, which of the following would best describe how you spend your time? Please assume	X	X			

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	that any other help you currently have would remain the same. (4 point scale) • I would be able to spend my time as I want, doing things I value or enjoy • I wouldn't do anything I value or enjoy with my time					
	Which of these statements best describes how having help to do things makes you think about feel about yourself? (4 point scale) • Having help makes me think and feel better about myself • Having help completely undermines the way I think and feel about myself	X			X	
	Thinking about the way you are helped and treated, and how that makes you think and feel about yourself, which of these statements best describes your situation? (4 point scale) • The way I'm helped and treated makes me think and feel better about myself • The way I'm helped and treated completely undermines the way I think and feel about myself	X			X	
Capability-Based Questionnaire for Assessing Well-Being in	Being loved and having friendship. (four point scale) • I am able to have a lot of love and contact with friends or family		X			

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
Patients with Chronic Pain	• I am not able to have any love or contact with friends or family					
	Being able to do things for fun (Including being in the 'right frame of mind' to experience enjoyment in life). (four point scale) • I am able to get a lot of enjoyment in life • I am not able to get any enjoyment in life		X			
	Being open, honest, believed, respected and understood (Not being defined as a person with ill health). (four point scale) • I am able to feel totally respected and positive about who I am • I am not able to feel respected or positive about who I am		X			
	Doing things which are productive and interesting (for example, hobbies, work or sport). (four point scale) • I am able to be totally active (both body and mind) • I am not able to be active (both body and mind)		X			
	Being independent and being able to make decisions. • I am able to be as independent as I want to be • I am not be able to be independent		X			
	The impact that I have on the lives of others (including: partners, family, colleagues and my local community). (four point scale)		X			

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• I am able to contribute to the well-being and happiness of those closest to me, in many ways • I am not at all able to contribute to the well-being and happiness of those closest to me					
	My health and mental well-being (including: pain, depression, sleep, mobility, medication side-effects). (four point scale) • I have no problems with my physical health or mental well-being • I have severe problems with my physical health or mental well-being	X		X		
	Knowing that in the future I will be able to cope. (four point scale) • When I think about the future I am able to feel completely confident that I will cope • When I think about the future I am not able to have any confidence that I will cope	X	X	X		
Child – and parent report questionnaire to explore capability of deaf children wearing a cochlear implant**T†	I understand the teacher, when she or he explains something. • Very true • A bit true • Not true ➔ If very true or a bit true: Is it difficult for you to understand what the teacher says? • For me it is easy to understand what the teacher says.		X		X	

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• For me it is a bit difficult to understand what the teacher says. • For me it is very complicated to understand what the teacher says. ➔ If easy to understand, then go to the next item. ➔ If a bit difficult or very difficult, go to the following question: Imagine that you would like to understand what the teacher says. Would you then be able to? • I would easily understand what the teacher says. • I would have difficulties understanding what the teacher says. • I would not understand what the teacher says. If would easily understand: How much would you like to understand what the teacher is saying? (3 point scale) • I would like it very much to understand what the teacher is saying. • I would not like to understand what the teacher is saying If would have difficulties, repeat the last question, plus: Why is it difficult for you to understand what the teacher says? (open question)					

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	If would not understand, repeat question about how much the child would like to understand the teacher, plus: Why can you not understand what the teacher says? (open question)					
	I watch television without subtitles. (3 point scale) • Very true • Not true	X	X	X	X	
	I have a pleasant relationship with my parents. (3 point scale) • Very true • Not true	X	X	X	X	
	Outside of school I meet with friends. (3 point scale) • Very true • Not true	X	X	X	X	
	I understand the feelings of other children, for example sadness or jealousy. (3 point scale) • Very true • Not true	X	X	X	X	
	I ask question to people I do not know, for example when I need something. (3 point scale) • Very true • Not true	X	X	X	X	
	I stand up for myself, for example when someone annoys me. (3 point scale) • Very true • Not true	X	X	X	X	
	I go somewhere by myself, for example to the city or the supermarket. (3 point scale)	X	X	X	X	

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• Very true • Not true					
	I often have fun. (3 point scale) • Very true • Not true	X	X	X	X	
Diabetes specific instrument for measuring patient reported outcomes and experiences in the Swedish National Diabetes Register *	How have you felt in general in the past four weeks?				X	
	How have you slept in the past four weeks?				X	
	Have you felt depressed in the past four weeks?				X	
	Has having diabetes been difficult in the past four weeks?				X	
	How have you dealing with your diabetes in the past four weeks?	X			X	
	Do you worry about getting too low blood sugar?	X			X	
	Do you worry that your blood sugar is too high?	X			X	
	Do you worry that your diabetes can cause other diseases or injuries?	X			X	
	Do you think your knowledge is sufficient to care for your diabetes?	X				
	How do you deal with your diabetes on a day-to-day basis?	X				
	How do you deal with your diabetes when your ordinary routines are difficult to follow?	X			X	
	How do you manage to eat in a way that you believe is good for you?					X
	How well are you able to stay as physically active as you believe is good for you?		X			

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	Does your diabetes prevent you from doing what you want?	X				
	Does your diabetes pose as an obstacle to spending time with your family, friends and others?	X	X			
	How often does low blood sugar prevent you from doing what you want?	X				
	How often does high blood sugar prevent you from doing what you want to?	X				
	How often are you prevented from doing what you want because your blood sugar fluctuates between high and low levels?	X				
	How helpful is the support for your diabetes care from family, friends and others close to you?	X	X			
	How well do other persons that you meet in your daily life support you in dealing with diabetes?	X	X			
	How well do other people who also have diabetes support you in dealing with your diabetes?	X	X			
	Do you get the support you need from your diabetes care provider?	X	X			
	Is it easy to contact your diabetes care provider when you need help with your diabetes?	X	X			
	Are you able to see a nurse as often as you feel is necessary for your diabetes?		X			
	Are you able to make visits with your nurse that fit your schedule?		X			
	Are you able to see the same nurse for your diabetes at every visit?		X			
	Are you able to see a doctor as often as you feel is necessary for your diabetes?		X			

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	Are you able to make visits with your doctor that fit your schedule?		X			
	Are you able to see the same doctor for your diabetes at every visit?		X			
	Are you able to talk about matters that are important to you at the appointments about your diabetes?		X			
	How satisfied are you with the medical devices available for you to monitor your blood sugar level?					X
	How satisfied are you with the medical devices that you have available for you to take insulin (for example, an insulin pen or insulin pump)?					X
	How satisfied are you with your medication treatment? The question includes all your medications you take.					X
ICECAP-A	Feeling settled and secure (4 point scale) • I am able to feel settled and secure in all areas of my life • I am unable to feel settled and secure in any areas of my life		X			
	Love, friendship and support (4 point scale) • I can have a lot of love, friendship and support • I cannot have any love, friendship and support	X	X			
	Being independent (4 point scale) • I am able to be completely independent • I am unable to be at all independent		X			
	Achievement and progress (4 point scale) • I can achieve and progress in all aspects of my life		X			

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• I cannot achieve and progress in any aspects of my life					
	Enjoyment and pleasure (4 point scale) • I can have a lot of enjoyment and pleasure • I cannot have any enjoyment and pleasure		X			
ICECAP-SCM	Having a say – Your ability to influence where you would like to live or be cared for, the kind of treatment you receive, the people who care for you (4 point scale) • I am able to make decisions that I need to make about my life and care most of the time • I am never able to make decisions that I need to make about my life and care			X		
	Being with people who care about you – Being with family, friends or caring professionals (4 point scale) • If I want to, I am able to be with people who care about me most of the time • If I want to, I am never able to be with people who care about me		X			
	Physical suffering – Experiencing pain or physical discomfort which interferes with your daily activities (4 point scale) • I rarely experience significant physical discomfort • I always experience significant physical discomfort				X	
	Emotional suffering – Experiencing worry or distress, feeling like a burden (4 point scale)				X	

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• I rarely experience emotional suffering • I always experience emotional suffering					
	Dignity – Being treated with respect, being spoken to with respect, having your religious or spiritual beliefs respected, being able to be yourself, being clean, having privacy (4 point scale) • I am able to maintain my dignity and self-respect most of the time • I am never able to maintain my dignity and self-respect		X			
	Being supported – Having help and support (4 point scale) • I am able to have the help and support that I need most of the time • I am never able to have the help and support that I need	X	X			
	Being prepared – Having financial affairs in order, having your funeral planned, saying goodbye to family and friends, resolving things that are important to you, having treatment preferences in writing or making a living will (4 point scale) • I have had the opportunity to make most of the preparations I want to make • I have not had the opportunity to make any of the preparations I want to make		X			
ICECAP-O	Love and Friendship (4 point scale) • I can have all of the love and friendship that I want • I cannot have any of the love and friendship that I want		X			

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	Thinking about the future(4 point scale) • I can think about the future without any concern • I can only think about the future with a lot of concern		X			
	Doing things that make you feel valued (4 point scale) • I am able to do all of the things that make me feel valued • I am unable to do any of the things that make me feel valued		X			
	Enjoyment and pleasure (4 point scale) • I can have all of the enjoyment and pleasure that I want • I cannot have any of the enjoyment and pleasure that I want		X			
	Independence (4 point scale) • I am able to be completely independent • I am unable to be at all independent		X			
OCAP-18	Until what age do you expect to live, given your family history, dietary habits, lifestyle and health status? • [Number to be entered]		X			
	Does your health in any way limit your daily activities, compared to most people of your age? • Yes • No	X	X			
	How suitable or unsuitable is your accommodation for your current needs? (5 point scale)					X

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• Very suitable • Very unsuitable					
	Please indicate how safe you feel walking alone in the area near your home? (5 point scale) • Very safe • Very unsafe				X	
	Please indicate how likely you believe it to be that you will be assaulted in the future (including sexual and domestic assault)? (5 point scale) • Very unlikely • Very likely					X
	I am able to express my views, including political and religious views. (5 point scale) • Strongly agree • Strongly disagree		X			
	I am free to use my imagination and to express myself creatively (e.g. through art, literature, music etc.). (5 point scale) • Strongly agree • Strongly disagree		X			
	At present how easy or difficult do you find it to enjoy the love, care and support of your family and friends? (5 point scale) • Very easy • Very difficult	X	X			
	In the past 4 weeks, how often have you lost sleep over worry? (5 point scale) • Never • Always				X	

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	I am free to decide for myself how to live my life. (5 point scale) • Strongly agree • Strongly disagree			X		
	I am able to respect, value and appreciate people around me. (5 point scale) • Strongly agree • Strongly disagree		X			
	Are you able to meet socially with friends, relatives or work colleagues? • Yes • No		X			
	Outside of any employment, in your everyday life, how likely do you think it is that you will experience discrimination? (5 point scale) • Very unlikely • Very likely				X	
	I am able to appreciate and value plants, animals and the world of nature (5 point scale) • Strongly agree • Strongly disagree		X			
	In the past 4 weeks, how often have you been able to enjoy your recreational activities? (5 point scale) • Always • Never		X			
	I am able to influence decisions affecting my local area. (5 point scale) • Strongly agree • Strongly disagree		X			
	[If the participants indicates that he or she has not bought a home] Which of these applies to	X				

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	your home? For which of the following reasons, if any, have you NOT bought your home? • I cannot afford to buy • I cannot obtain a mortgage • I think it is a bad time to buy • There is a lack of available housing to buy • Some other reason					
	In your current or future employment, how likely do you think it is that you will experience discrimination? (5 point scale) • Very likely • Very unlikely				X	
OxCAP-MH	Does your health in any way limit your daily activities, compared to most people of your age? (5 point scale) • Never • Always	X	X			
	Are you able to meet socially with friends or relatives? (5 point scale) • Always • Never		X			
	In the past 4 weeks, how often have you lost sleep over worry? (5 point scale) • Always • Never				X	
	In the past 4 weeks, how often have you been able to enjoy your recreational activities? (5 point scale) • Always		X			

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• Never					
	How suitable or unsuitable is your accommodation for your current needs? (5 point scale) • Very suitable • Very unsuitable					X
	Please indicate how safe you feel walking alone in the area near your home: (5 point scale) • Very safe • Very unsafe				X	
	Please indicate how likely you believe it to be that you will be assaulted in the future (including sexual and domestic assault): (5 point scale) • Very unlikely • Very likely				X	
	How likely do you think it is that you will experience discrimination? (5 point scale) • Very unlikely • Very likely				X	
	On what grounds do you think it is likely that you will be discriminated against? • Race/ethnicity • Gender • Religion • Sexual orientation • Age • Health or disability (incl. mental health)	X				
	I am able to influence decisions affecting my local area. (5 point scale)		X			

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• Strongly agree • Strongly disagree					
	I am free to express my views, including political and religious views. (5 point scale) • Strongly agree • Strongly disagree		X			
	I am able to appreciate and value plants, animals and the world of nature. (5 point scale) • Strongly agree • Strongly disagree		X			
	I am able to respect, value and appreciate people around me. (5 point scale) • Strongly agree • Strongly disagree		X			
	I find it easy to enjoy the love, care and support of my family and/or friends. (5 point scale) • Strongly agree • Strongly disagree	X	X			
	I am free to decide for myself how to live my life. (5 point scale) • Strongly agree • Strongly disagree			X		
	I am free to use my imagination and to express myself creatively (e.g. through art, literature, music, etc.). (5 point scale) • Strongly agree • Strongly disagree		X			
	I have access to interesting forms of activity (or employment). (5 point scale) • Strongly agree • Strongly disagree		X			

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
Women's capability index**	Physical strength.	X	X			X
	During the last week, did you have any physical health problems? • No • Yes Malaria • Yes Diarrhea • Yes Fever / Headache / Backache / periods Yes, other (specify)				X	
	If Yes, How much were your activities affected by these physical problems? (5 point scale) • Not at all • Could not do physical activities	X	X			
	Inner wellbeing	X	X	X		X
	During the last week, did you have any emotional worry or problem? • Yes • No				X	
	If yes, did these emotional problems keep you from doing your usual daily activities (working in the garden, household chores)? • Not at all • Yes, Could not do activity	X	X			
	Household wellbeing	X	X		X	X
	Does your household have a toilet? • Yes, flush toilets • Yes, pit latrines				X	

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• Yes, VIP latrines • Neighbor's toilet, because *specify…* • Bush, because *specify…*					
	Are you able to take good care of your household members, as you wish, such as bathing, washing? • Yes • No, because I don't have enough money • No, because I don't have enough time • No, because… *specify*	X	X			
	Have you ever been beaten by your husband (or ex husband) or by other household member? (4 point scale) • Yes often • Never				X	
	Community relations	X	X		X	X
	How safe do you feel walking alone in your village when it is getting dark? (5 point scale) • Completely safe • Not at all safe				X	
	To what extent do you feel that people in your community treat you with respect? (5 point scale) • Entirely • Not at all				X	
	Economic security	X	X		X	X
	What do you do for a living? • Farming • Business (trade) • Private sector (NGOs, CBOs, Farm estate)				X	

Instrument	Questions	Content related to elements that support or burden access to options	Content related to options themselves	Content which directly assesses the perception of freedom	Content assessing functioning	Other content
	• Artisan (Tailor, builder, mechanic) • Domestic work • Other					
	Imagine a crisis such as your crops fail, how confident are you that you can feed your family for 4 weeks? (5 point scale) • Very confident • No confident at all	X	X			
	Happiness				X	
	Taking all things together in your life, how satisfied are you with your life? (5 point scale) • Completely satisfied • Not satisfied at all / very unsatisfied				X	
	Taking all things together in your life, would you say you are… (4 point scale) • Very happy • Not at all happy				X	

* The authors were unable to identify the response options for this questionnaire.

** Due to the length of the questionnaire, the authors decided to present some exemplary questions per domain. The row which mentions the domain shows which elements of option freedom, functioning and other aspects are covered by the content of that specific domain.

ᵀ Questions have been translated from Dutch into English by me. Please note that these items are therefore not an officially validated translation.

† Due to the complex adaptive question methodology used in this study, we decided to present only one item completely. Other items follow a similar pattern: first asking about a functioning, followed by reflecting on that functioning in order to assess whether that level of functioning is a choice or a consequence of limited capabilities.

Table has been published previously in Ubels et al. (2022b).

10.2 Appendix to Chapter 3. ENTREQ-checklist

Appendix Table 3. ENTREQ-checklist

No	Item	Page were item is addressed
1	Aim	60
2	Synthesis methodology	61
3	Approach to searching	61
4	Inclusion criteria	61
5	Data sources	61
6	Electronic search strategy	35
7	Study screening methods	35
8	Study characteristics	41, 157
9	Study selection results	36, 61
10	Rationale for appraisal	61
11	Appraisal items	61
12	Appraisal process	61
13	Appraisal results	62
14	Data extraction	61
15	Software	61
16	Number of reviewers	62
17	Coding	61
18	Study comparison	61
19	Derivation of themes	61
20	Quotations	Throughout results section of Chapter 3 (Section 3.3)
21	Synthesis output	Throughout results section of Chapter 3 (Section 3.3)

10.3 Appendix to Chapter 4

10.3.1 Edits to the MIC database

The MIC research team performed a number of edits to the data. In edit (1) responses from participants that finished the survey in less time than 20 minutes were eliminated. Responses from participants who completed the survey in a time between 20 to 25 minutes were subject to closer inspection. For edit (2) the mobility item of the EQ-5D-5L was asked

twice at different places in the survey. The responses between the items that varied by more than one point on the Likert scale were deleted. Responses that varied by one point on the Likert scale were noted for further inspection. In edit (3), the responses to item one of the SF-36 V2 and an item where participants rate their own health were compared. Responses that differed by more than one point on the Likert scale were deleted. Responses that differed by one point on the Likert scale were noted for further inspection. In edit (4), the first item of the SF-36 V2 and item 9a from the QWB-SA were identical, Responses with a difference in score of more than one on the Likert scale between these items were removed. Responses that differed only one point on the Likert scale were noted for further scrutiny. Edit (5) consisted of a similar comparison between question 9A and the “own health” item. Edit (6) was based on a comparison between item 4 of the EQ-5D-5L and item 22 of the AQoL-8D. Both of these items cover the experience of pain. Responses were removed when the responses between these items differed by two or more points on the Likert scale. In edits (7) and (8) the responses that were marked for further scrutiny based on edits (1), (2), (3), (4), (5), and (6) were analyzed. In the case of edit (7), the inconstancies from edits (2), (3), (4), (5), and (6) were summed. Responses were removed when there were more than two inconsistencies and the time of finishing the survey was less than 25 minutes. In edit number (8), responses that had more than three inconstancies were removed.

10.3.2 The selected items and their link to the qualitative framework.

Appendix Table 4. Original themes, quotes related to those themes, items from the multi instrument comparison database with their code and the original instrument to which these items belong.

Theme	Subtheme	Illustrative quote	Items MIC* database	MIC code †	Instrument ‡
Perceived access to options	Items directly linked to theme		During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours or group? Not at all	sf20	SF-36 V2
			During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)? All of the time	sf32	SF-36 V2
	Access due to physical wellbeing	*[The chest infection] just made it miserable for a week or two, I couldn't get out or about…[Male, 75], Al-Janabi et al. (2012)*	Do I need any help looking after myself? I need no help at all	aqol1_4D	AQOL-4D
			When doing household tasks: (For example: preparing food, gardening, using the video recorder, radio, telephone or washing the car.) I need no help at all	aqol2_4D	AQOL-4D
			Thinking about your health and your role in your community (that is to say neighbourhood, sporting, work, church or cultural groups): My role in the community is unaffected by my health	aqol4	AQOL-8D
			Thinking about your health and your relationship with your family: My role in my family is unaffected by my health	aqol9	AQOL-8D
			Thinking about washing yourself, toileting, dressing, eating or looking after your appearance: These tasks are very easy for me	aqol19	AQOL-8D
			How often does pain interfere with your usual activities? Never	aqol24	AQOL-8D

			How much help do you need with tasks around the house (e.g., preparing food, cleaning the house or gardening): I can do all these tasks very quickly and efficiently without any help	aqol30	AQOL-8D
			During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your PHYSICAL health? Cut down the amount of time you spent on work or other activities. All of the time	sf13	SF-36 V2
			During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your PHYSICAL health? accomplished less than you would like All of the time	sf14	SF-36 V2
			During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your PHYSICAL health? were limited in the kind of work or other activities All of the time	sf15	SF-36 V2
			During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your PHYSICAL health? had difficulty performing work or other activities (for example, it took extra effort) All of the time	sf16	SF-36 V2
			During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? Not at all	sf22	SF-36 V2

	Access due to emotional wellbeing	*I started getting depression… it's like yesterday, I didn't have a wash, I didn't have a shave, I didn't get up, I didn't even unlock the door, and that was it (Male, not employed, A), Kinghorn et al. (2015)*	During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your EMOTIONAL problems (such as feeling depressed or anxious)? Cut down the amount of time you spent on work or other activities All of the time.	sf17	SF-36 V2
			During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your EMOTIONAL problems (such as feeling depressed or anxious)? accomplished less than you would like All of the time	sf18	SF-36 V2
			During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your EMOTIONAL problems (such as feeling depressed or anxious)? didn't do work or other activities as carefully as usual All of the time	sf19	SF-36 V2
	Access due to social wellbeing	*When I came out of hospital he [husband] done everything I mean, he cooked the food and he's never cooked in his life [laugh] … And all the washing, ironing he did. (Female, 72 years, PC), Sutton and Coast (2014)*	No relevant items		
	Access due to environmental wellbeing	*"I went in for this flat because it's wheelchair friendly … I'm hoping that I'd lay here in a box, because it was a very deliberate act of me to look for somewhere where I can be independent for as long as possible." (Female, 67 years, GP), Sutton and Coast (2014)*	Thinking about how easy or difficult it is for you to get around by yourself outside your house (eg shopping, visiting): Getting around is enjoyable and easy	aqol3	AQOL-8D
			Thinking about how easily I can get around my home and community: I get around my home and community by myself without any difficulty	aqol3_4d	AQOL-4D

	Access due to activity wellbeing	*"Because of the arthritis and that, I can't work, so there's… work mates, you know, no Friday night when you've got the wages and you can enjoy it, a couple of beers. There's none of that." (Male, Not employed), Kinghorn et al. (2015)*	No relevant items		
	Access due to finances	*[Regarding the choice to do a NIPT] "If I had to pay for it, I would borrow from my friends or relatives. But I would just do anything possible to avoid a miscarriage." Kibel and Vanstone (2017)*	No relevant items		
		"I'm reasonably fortunate… in so far as that we've got the two pensions… we're able to go off... We grabbed a cheapie flight at the end of April... flew down to Nice..." (male, aged 70), Grewal et al. (2006)			
	Access due to technology	*"I think the insulin pump is fantastic. Because it gives me freedom." (#24; Woman, 64 years old, Type 1 DM), Engström et al. (2016)*	Thinking about your mobility, including using any aids or equipment such as wheelchairs, frames, sticks: I am very mobile	aqol15	AQOL-8D
Perceived control	Management	*"This morning, I got up—5 o'clock—I took my first pain killers, went back to bed again so that I was ready to get up to have my shower at half past six, or else, by the time you start taking them they haven't taken effect and you're trying to move around. So, yeah, you've got to think ahead…" (Female, not employed, A), Kinghorn et al. (2015)*	How much do you feel you can cope with life's problems? Completely	aqol21	AQOL-8D

		[Regarding the management of diabetes] "It's not easy, it's an endless struggle to try to maintain good blood glucose levels. (...) It's like walking a line." (#24; Woman, 64 years old, Type 1 DM), Engström et al. (2016)			
	Evaluation	*It is a constant sadness, that I've lost my sight... (...) But it's nothing I get hung up on in my everyday life. (...) I consider myself as having a good quality of life. (#1; Man, 49 years old, Type 1 DM)*	And still thinking about the last seven days: how often did you feel worried: Never	aqol18	
			How much of a burden do you feel you are to other people? Not at all	aqol26	
			How often do you feel in control of your life? Always	aqol29	
Option wellbeing	Item directly linked to theme		I am satisfied with my life Strongly agree	SWLS_c	SWLS
			Overall, how satisfied are you with your life nowadays	ONSi	ONS
			Thinking about your own life and personal circumstances, how satisfied are you with your life as a whole?	PWI_a	PWI
	Physical wellbeing	*"I just wouldn't want to be in pain all the time." (Female, 72 years, PC), Sutton and Coast (2014)*	How satisfied are you with your health? Completely satisfied	PWI_c	PWI
			In general, would you say your health is Excellent	sf1	SF-36 V2
	Emotional wellbeing	*It's sad not daring to go [on a trip]. (...) Since it [hypo- glycaemia] is a threat, it feels like a lower quality of life. (...) You get a little scared of exposing yourself to situations other than what you are used to. (#17; Woman, 60 years old, Type 2 DM), Engström et al. (2016)*	How often do you feel sad? Never	aqol5	AQOL-8D
			How often do you feel happy? All the time	aqol20	AQOL-8D
			How often do you feel pleasure? Always	aqol25	AQOL-8D
			How content are you with your life? Extremely	aqol27	AQOL-8D
			How often do you feel depressed? Never	aqol33	AQOL-8D

			How often did you feel in despair over the last seven days? Never	aqol35	AQOL-8D
			Thinking about how I generally feel: I do not feel anxious, worried or depressed	aqol11_4D	AQOL-4D
			Enjoyment and pleasure I can have a lot of enjoyment and pleasure	ic05	ICECAP - A
			Overall, how happy did you feel yesterday?	ONSk	ONS
			Overall, how anxious did you feel yesterday?	ONSl	SF-36 V2
			How much of the time during the past 4 weeks … Have you been a very nervous person All the time	sf24	SF-36 V2
			How much of the time during the past 4 weeks … Have you felt so down in the dumps that nothing could cheer you up All the time	sf25	SF-36 V2
			How much of the time during the past 4 weeks … Have you felt down All the time	sf28	SF-36 V2
			How much of the time during the past 4 weeks … Did you feel worn out All the time	sf29	SF-36 V2
			How much of the time during the past 4 weeks … Have you been a happy person All the time	sf30	SF-36 V2
	Social wellbeing	*"One should take good care of the kids and the entire family, so that everyone is healthy and they can work properly and prosper." Greco et al. (2015)*	Your close relationships (family and friends) are: Very satisfying	aqol10	AQOL-8D
			How much do you enjoy your close relationships (family and friends)? Immensely	aqol23	AQOL-8D
			Your close and intimate relationships (including any sexual relationships) make you feel: Very happy	aqol34	AQOL-8D

			Thinking about my relationship with other people: I have plenty of friends, and am never lonely	aqol5_4D	AQOL-4D
			How satisfied are you with your personal relationships? Completely satisfied	PWI_e	PWI
	Environmental wellbeing	*"A house should have a toilet, a bathing shelter, there should be a rubbish pit, and the house should be well taken care of. Even if you have all these things but they are not put to good use, diseases will be there." Greco et al. (2015)*	How satisfied are you with feeling part of your community? Completely satisfied	PWI_g	PWI
	Activity wellbeing	*"Work is important. Just to go out and do things that aren't mind numbing if you know what I mean." (F employed), Kinghorn et al. (2015)*	No relevant items		
Self-realization	Having a role	*"I do like playing…competitive sport…it's got a bit of an edge …. I suppose through that there's a bit of an achievement thing and it's quite nice to be in a team or to be a captain for one of the teams" (Male, 29), Al-Janabi et al. (2012)*	Achievement and progress I can achieve and progress in all aspects of my life	ic04	ICECAP - A
			Overall, to what extent do you feel that the things you do in your life are worthwhile? Completely worthwhile	ONSj	ONS
			How satisfied are you with what you are achieving in life? Completely satisfied	PWI_d	PWI
	Having dignity	*"A person who changes clothes is seen as living a good life. She changes dirty clothes after a bath, and puts on clean ones, and looks good. When she is amongst people, she is not shy. As for me, I may have to wash the few I have to put on when I go in public." Greco et al. (2015)*	No relevant items		

	Being independent	*"I would like to drive a bit more. Because I'm losing my independence. I have to rely on my husband to take me shopping now." (Female, Retired), Kinghorn et al. (2015)*	Being independent I am able to be completely independent	ic03	ICECAP - A
		"A person should be independent because when sick she doesn't wait for someone to tell her what to do, men at times neglect that you are struggling." Greco et al. (2015)			
	Self-determination	*[Regarding the choice of conducting NIPT] "I just really think that women should be given ownership of the information and they can decide what they want to do." Kibel and Vanstone (2017)*	In most ways my life is close to my ideal Strongly Agree	SWLS_a	SWLS
			So far I have gotten the important things I want in life Strongly Agree	SWLS_d	SWLS

* MIC: Multi Instrument Comparison, from the Multi Instrument Comparison Study (Richardson et al., 2012).
† Codes of the items in the MIC database.
‡ Abbreviations of original items of instruments: AQOL-4D (-8D): Assessment of Quality of Life instrument (-4D: four dimensions, -8D: eight dimensions), ICECAP-A: ICEpop CAPability measure for Adults, ONS: Integrated household survey from the Office of National Statistics, PWI: Personal Wellbeing Index, SF-36 V2: 36-Item Short Form Health Survey Version 2, SWLS: Satisfaction With Life Survey.
Table has been published previously in Ubels et al. (2022c).

10.3.3 Theoretical adjustments to the measurement model

The fit indices indicated inadequate model fit of the first measurement model (see Model 1 in Table 5). Several theoretical adjustments were therefore made to test whether model fit could be improved. One change was that the items sf1 and PWI_c (see for items associated with these codes Appendix Table 4) were both cross-loaded on the “Perceived Access to Options“ construct. This was done because both items reflect the general satisfaction that individuals experience with their health and the construct “Perceived Access to Options” reflects the health-related capabilities of individuals.

Further theoretical changes were made with the items sf13, sf14, sf15, and sf16. Based on the layout of the MIC survey, I hypothesized that these items share covariance since they were presented as a testlet. I linked these items to an additional orthogonal factor to account for this covariance. A similar orthogonal factor was constructed for the items sf17, sf18, and sf19. The model fit of the measurement model after these theory-based changes can be found in Table 5 under Model 2.

10.3.4 Data-driven adjustments to the measurement model

Model fit was still inadequate after theoretical adjustments (see Model 2 in Table 5). I hypothesized that one source of this misfit was unaccounted for covariance between items that showed similarities. Furthermore, the theoretical model itself could be miss-specified, which meant that the items reflect different constructs than I hypothesized based on the theoretical framework developed in Chapter 3. I decided to inspect residual correlations and modification indices to study whether the theoretical framework needed to be improved and to identify further sources of misfit.

The first and largest change concerns the restructuring of two constructs. Based on modification indices and residual correlations I identified that two constructs were miss-specified: “Option Wellbeing” and “Self-Realization”. These constructs respectively reflect the emotional aspects of wellbeing, such as satisfaction and happiness or sadness, and wellbeing derived from living a meaningful life. Both these constructs thus reflect the experienced wellbeing of an individual. The residual correlations and the modification indices indicated however that model fit could be improved by

restructuring these two constructs into two new constructs: “Reflective Wellbeing” and “Affective Wellbeing”. The “Reflective Wellbeing” construct reflects both life satisfaction and the experience of having a meaningful life. These are cognitive appraisals of wellbeing. The “Affective Wellbeing” construct reflects different emotional experiences related to wellbeing, such as happiness, sadness, and anxiety. These adjustments resulted in an improvement in model fit.

The second group of changes concerns changes to individual items to improve model fit. Items aqol23, aqol27, and sf29, which load on the factor “Affective Wellbeing”, were cross-loaded on the factor “Reflective Wellbeing”. The item aqol26, which loaded on the factor “Perceived Control” was also allowed to cross-load on the factor “Perceived Access to Options”. The items aqol5_4D and aqol10, which initially loaded on the “Self-Realization” factor, were loaded on the factor “Affective Wellbeing”. The item ONSk was removed from the measurement model. This item covers the experience of happiness. However, its factor loadings indicated that the item was more closely linked to the “Reflective Wellbeing” scale (standardized loading of 0.660) than the “Affective Wellbeing” factor (standardized loading of 0.185). This might have been a consequence of the layout of the MIC survey questionnaire. Two other ONS items (ONSi and ONSj, reflecting life satisfaction and experiencing meaningfulness respectively), which both also show high loadings on the “Reflective Wellbeing” factor, might have influenced the response on the ONSk item since the ONSk item followed ONSi and ONSj. As such, it seems that the content of the item did not reflect the actual experience of an individual. As a consequence, I decided to delete the item ONSk.

A considerable difference between the measurement model that was based on the theoretical framework and the measurement model that was developed after exploratory studying the data concern three items from the ICECAP-A: item ic03 (covers the capability to be independent), ic04 (covers the capability to achieve and progress, and ic05 (covers the capability to enjoy and experience pleasure). Items ic03 and ic04 initially loaded on the construct “Self-Realization”. Item ic05 initially loaded on the “Option Wellbeing” construct.

Residual correlations indicated however that model fit could be improved by loading ic03 on the factors “Perceived Control” and “Perceived Access to Options”. The latter

factor loading indicates that this item reflects the health-related capabilities of an individual instead of their experienced “Self-Realization”. Also, item ic05 loaded on the factor “Perceived Control”, as well as the newly formed factor “Affective Wellbeing”. Lastly, residual correlations indicated that item ic04 loaded more appropriately on the “Perceived Control” factor. The fact that model fit could be improved by loading each of these items on the “Perceived Control” factor could be explained due to the similarities between the wording of these items (I am able to…, I can…) and the concept self-efficacy (Bandura, 2001; Frei et al., 2009). Self-efficacy represents an individual’s perceived ability to be able to do things (Bandura, 2001). Instruments measuring this concept use wording similar to the wording of the items of the ICECAP-A (Frei et al., 2009).

The third group of changes that require further discussion concern the seven specific factors that were created to account for covariances that are caused by similarities in the content of items and the layout of the MIC questionnaire. These seven factors adjust for the covariance of: (1) items that cover social aspects, (2) items that cover the need for help or support, (3) items that reflect happiness, (4) items that reflect anxiety, (5) the physical limitation testlet of the SF-36 V2, (6) the emotional limitation testlet of the SF-36 V2, (7) items from the SF-36 V2 that reflect negative affect, such as depression. The latter two factors were allowed to correlate with each other due to the similarities in wording between the items of the two factors, which means that these two specific factors are not orthogonal. The other five factors were however orthogonal.

Also, some error terms of pairs of items were correlated with each other. An example of such a pair was already given: PWI_c and sf1, which both cover the perceived general health of individuals. Besides this pair, also the error terms of the items aqol24 and sf22 were correlated with each other since both specifically inquire about the extent to which pain influences normal activities or work. Additionally, the error terms of the items sf20 and sf32 were correlated with each other, due to both inquiring whether health and emotional problems interfered with social activities. The last pair of items with correlated error terms were aqol2_4D and aqol30, due to them both inquiring about the need for help in household tasks. The correlation values of these pairs can be found in Chapter 4.

10.3.5 Item selection for the development of the WeRFree instrument

The measurement model that was developed to test the theoretical framework developed in Chapter 4 consisted of 56 items. Of these 56, 28 items remained after removing the items from the “Perceived Control” construct and items that cross-loaded on multiple factors. To develop a parsimonious instrument, these 28 items were further reduced in number by retaining one item from groups of items that cover similar content. This selection was based on the item difficulty properties of single items, their ceiling – and floor effects, as well as their content.

The following items covered the ability of individuals to take care of themselves: aqol1_4D, aqol3_4D, and aqol19. A study of the proportion of the response options for these items showed that 85% of the MIC study sample chose the optimal response option for the items aqol11_4D and aqol13_4D. In the case of item aqol19, 66% of the sample chose the optimal response option. In order to maximize the information derived from an item that represents the ability to take care of oneself, I decided to retain aqol19 for the WeRFree instrument.

Another group of items reflected the experience of happiness. These items have the codes ic05, sf30, aqol20, and aqol25. On the overall “Affective Wellbeing” construct, these items reflect the “positive” aspects of emotional wellbeing. Therefore, additional attention was paid to ceiling effects. The highest ceiling effect was identified in item ic05, where 34% of the sample responded to the highest response option. Also, the remaining three items showed a large clustering around their two highest response options (sf30: 69%, aqol20: 83%, and aqol25: 80%). From this set, I decided to keep sf30, due to it having the comparatively lowest clustering around the two highest response options.

There was also a group of three items that reflect anxiety: aqol11_4D, ONSl, and sf24. From these items, first I decided to remove the item ONSl because of its wording. This is the only item related to the “Affective Wellbeing” construct that uses the wording “yesterday” when asking if an individual experienced anxiety. Qualitative research has indicated that such wording influences responses to items (Sanghera et al., 2021). Since the other items did not include this wording and the findings of the study by (Sanghera et al., 2021), I decided to not include the item ONSl in the WeRFree

instrument, so that potential responses items are equally not influenced by specific recall periods. From the remaining items, I decided to keep item sf24. I made this choice because the clustering around the two highest most optimal response options (e.g., individuals are unaffected by anxiety) showed the lowest clustering in this item (sf24: 71%, aqol11_4D: 80%).

The items sf24, sf25, and sf28 from the SF-36 V2 also shared variance, since they together form an emotional wellbeing subscale. From these items, sf24 was already retained for the instrument as indicated in the last paragraph. For the choice between sf25 and sf28, I paid special attention to floor effects, since this group of items represent the “negative” aspects of emotional wellbeing. Between sf25 and sf28, I decided to keep sf25, because it had a lower floor effect associated with it (sf25: 2%, sf28: 4%).

Items aqol5_4D, aqol10, aqol23, aqol34, and PWI_e all shared that they covered different aspects of relationships. From this group, I kept the item PWI_e for two reasons. One, it had the most evenly spread of responses over its different response options. Two, it was the only item that followed a Likert scale format. The other items varied in their wording over different response options. An example of such variety can be found in the item aqol5_4D, which has the response option “having plenty of friends” and “feeling lonely”, which seem to reflect related but different constructs.

The last group of items was sf13, sf14, sf15, and sf16, which are part of a testlet in the SF-36 V2 that covers how much physical health problems affect various aspects of daily life, such as work or usual activities. Based on the items that were already included (aqol19 and aqol4), I believed these elements to be already sufficiently covered. Therefore, I did not include these four items in the WeRFree instrument.

Besides groups of items, there were also pairs of items that shared variance due to covering similar content. From these pairs, one item was selected to be retained for the WeRFree instrument. One pair of items cover the impact of pain on usual activities and work: aqol24 and sf22. From this pair, I decided to keep aqol24, since responses were more evenly spread over the different response options. The last pair of items were aqol2_4D and aqol30, which both specifically cover the need for support in doing

household tasks. From this pair, I retained the item aqol30 for the WeRFree scale, due to its lower ceiling effect (47% for aqol30, 70% for aqol2_4D).

10.3.6 The WeRFree instrument

Reflective Wellbeing

	Completely dissatisfied					Neutral					Completely satisfied
	0	1	2	3	4	5	6	7	8	9	10
Thinking about your own life and personal circumstances, how satisfied are you with your life as a whole?											
How satisfied are you with your personal relationships?											
How satisfied are you with feeling part of your community?											
	Not at all Worthwhile					Neutral					Completely worthwhile
Overall, to what extent do you feel that the things you do in your life are worthwhile?											

	Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree
In most ways my life is close to my ideal:							
So far I have gotten the important things I want in life:							

Affective Wellbeing

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
How much of the time during the past 4 weeks have you been a very nervous person?					
How much of the time during the past 4 weeks have you been a happy person?					

How often do you feel sad?
O Never
O Rarely
O Some of the time
O Usually
O Nearly all the time

How often did you feel in despair over the last seven days?
O Never
O Occasionally
O Sometimes
O Often
O All the time

Perceived Access to Options

Thinking about how easy or difficult it is for you to get around by yourself outside your house (eg shopping, visiting):
O Getting around is enjoyable and easy
O I have no difficulty getting around outside my house
O A little difficulty
O Moderate difficulty
O A lot of difficulty
O I cannot get around unless somebody is there to help me

Thinking about your health and your role in your community (that is to say neighborhood, sporting, work, church or cultural groups):
O My role in the community is unaffected by my health
O There are some parts of my community role I cannot carry out
O There are many parts of my community role I cannot carry out
O I cannot carry out any part of my community role

Thinking about washing yourself, toileting, dressing, eating or looking after your appearance:
O These tasks are very easy for me
O I have no real difficulty in carrying out these tasks
O I find some of these tasks difficult, but I manage to do them on my own
O Many of these tasks are difficult, and I need help to do them
O I cannot do these tasks by myself at all

How often does pain interfere with your usual activities?
O Never
O Almost never
O Sometimes
O Often
O Always

How much help do you need with tasks around the house (eg preparing food, cleaning the house or gardening):
O I can do all these tasks very quickly and efficiently without any help
O I can do these tasks relatively easily without help
O I can do all these tasks only very slowly without help
O I cannot do most of these tasks unless I have help
O I can do none of these tasks by myself

Appendix Table 5. Proportion of responses per response option of items in the WeRFree instrument

Items and response option level	Percentage of sample choosing specific response option
PWI_a	
1	1.9%
2	1.9%
3	4.6%
4	6.0%
5	6.3%
6	10.7%
7	10.3%
8	18.9%
9	23.3%
10	0.9%
11	5.3%
PWI_e	
1	2.9%
2	3.0%
3	3.9%
4	4.4%
5	4.7%
6	10.6%
7	07.5%
8	11.9%
9	16.6%
10	18.1%
11	16.4%
PWI_g	
1	2.5%
2	2.6%
3	3.7%
4	3.9%
5	5.0%
6	19.7%
7	10.0%
8	14.8%
9	17.7%
10	11.9%
11	7.9%
ONSj	
1	2.2%
2	3.1%
3	3.7%
4	3.8%
5	4.6%
6	12.2%
7	12.2%
8	16.2%
9	19.3%
10	14.6%
11	8.1%

Items and response option level	Percentage of sample choosing specific response option
SWLS_a	
1	6.7%
2	13.6%
3	13.0%
4	12.3%
5	25.3%
6	25.0%
7	4.1%
SWLS_d	
1	4.6%
2	8.5%
3	11.2%
4	12.0%
5	22.3%
6	30.7%
7	10.6%
aqol5	
1	3.8%
2	9.0%
3	40.0%
4	40.8%
5	6.4%
aqol35	
1	1.2%
2	7.0%
3	14.0%
4	22.0%
5	55.8%
sf24	
1	2.1%
2	7.5%
3	19.9%
4	31.3%
5	39.3%
sf30	
1	5.0%
2	15.7%
3	25.2%
4	43.6%
5	10.6%
aqol3	
1	1.3%
2	4.3%
3	8.2%
4	15.4%
5	34.9%
6	35.9%
aqol4	
1	5.1%
2	9.2%
3	18.4%

Items and response option level	Percentage of sample choosing specific response option
4	67.3%
aqol19	
1	0.1%
2	1.9%
3	11.4%
4	20.7%
5	65.8%
aqol24	
1	3.5%
2	10.5%
3	23.9%
4	33.2%
5	28.9%
aqol30	
1	0.8%
2	6.0%
3	15.0%
4	30.9%
5	47.3%

Table has been published previously in Ubels et al. (2022c).

Appendix Table 6. Item-rest correlations per scale

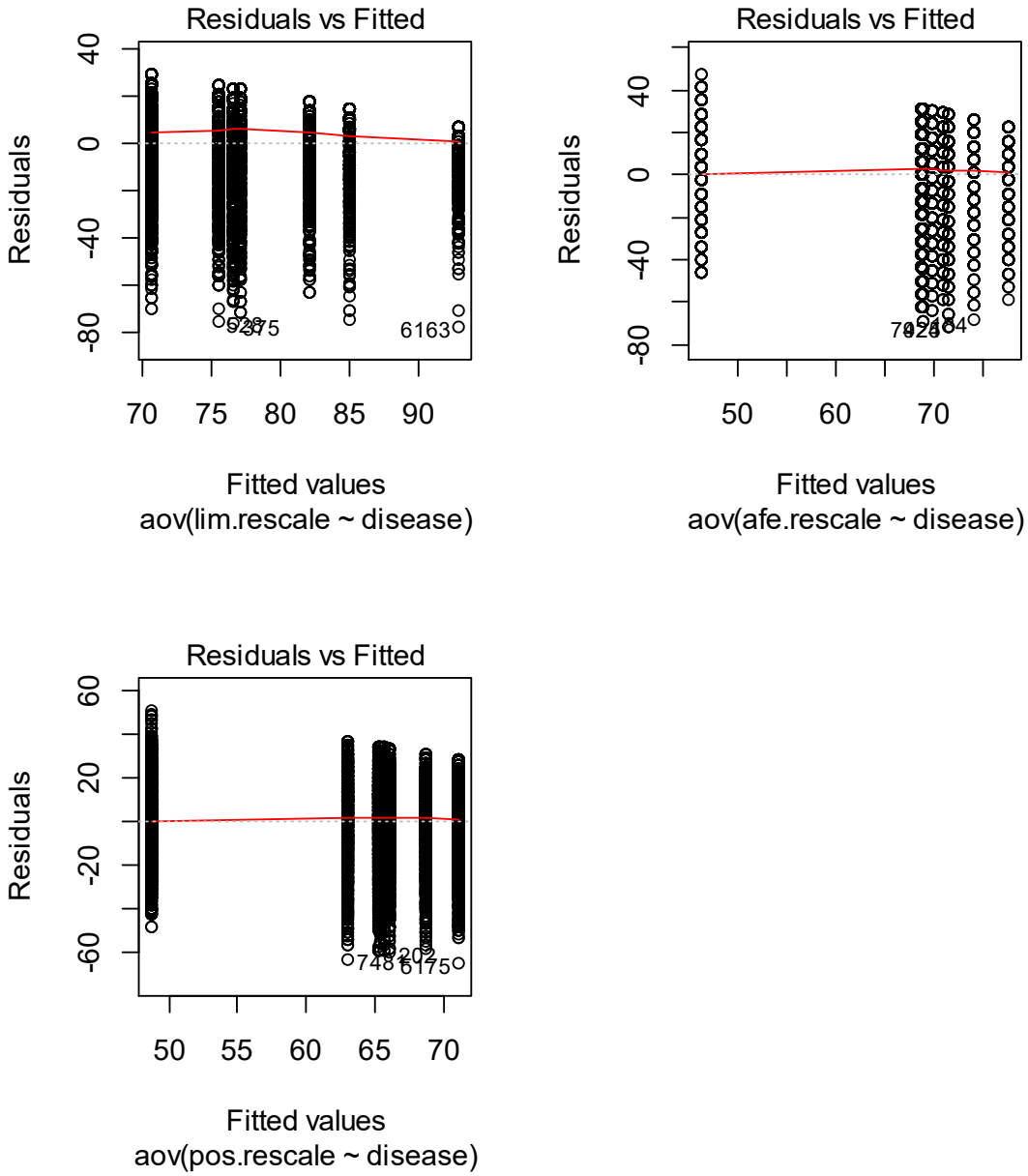
Item	Item-rest correlation
PWI_a	0.79
PWI_e	0.67
PWI_g	0.64
ONSj	0.77
SWLS_a	0.77
SWLS_d	0.72
sf24	0.58
sf30	0.62
aqol5	0.74
aqol35	0.73
aqo3	0.78
aqol4	0.70
aqol19	0.73
aqol24	0.70
aqol30	0.76

The items linked to the codes can be found in Appendix Table 4.
Table has been published previously in Ubels et al. (2022c).

10.4 Appendix to Chapter 5.

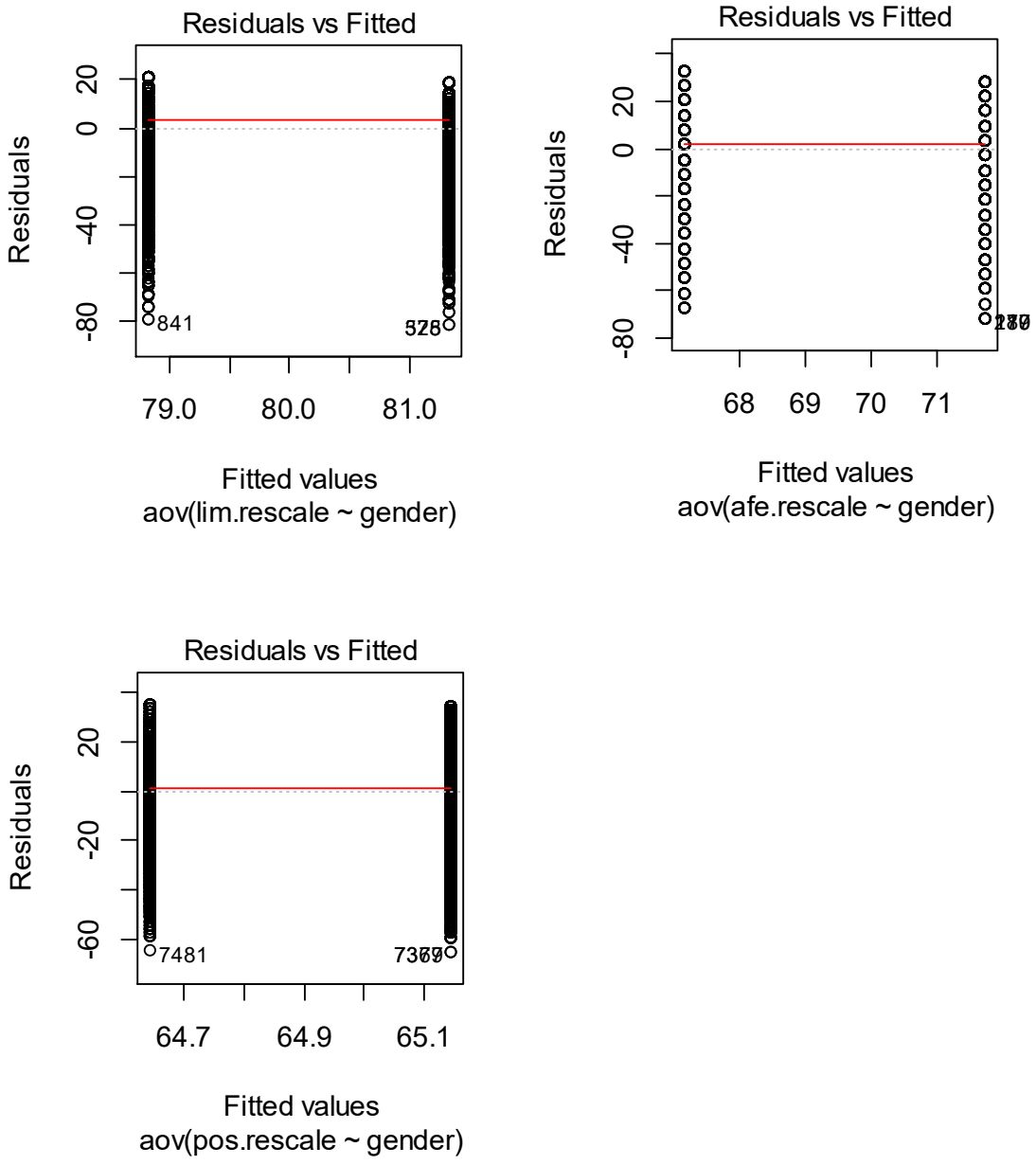
10.4.1 Residuals vs fitted values

Figure 4. Residual vs. fitted values for health condition subsample



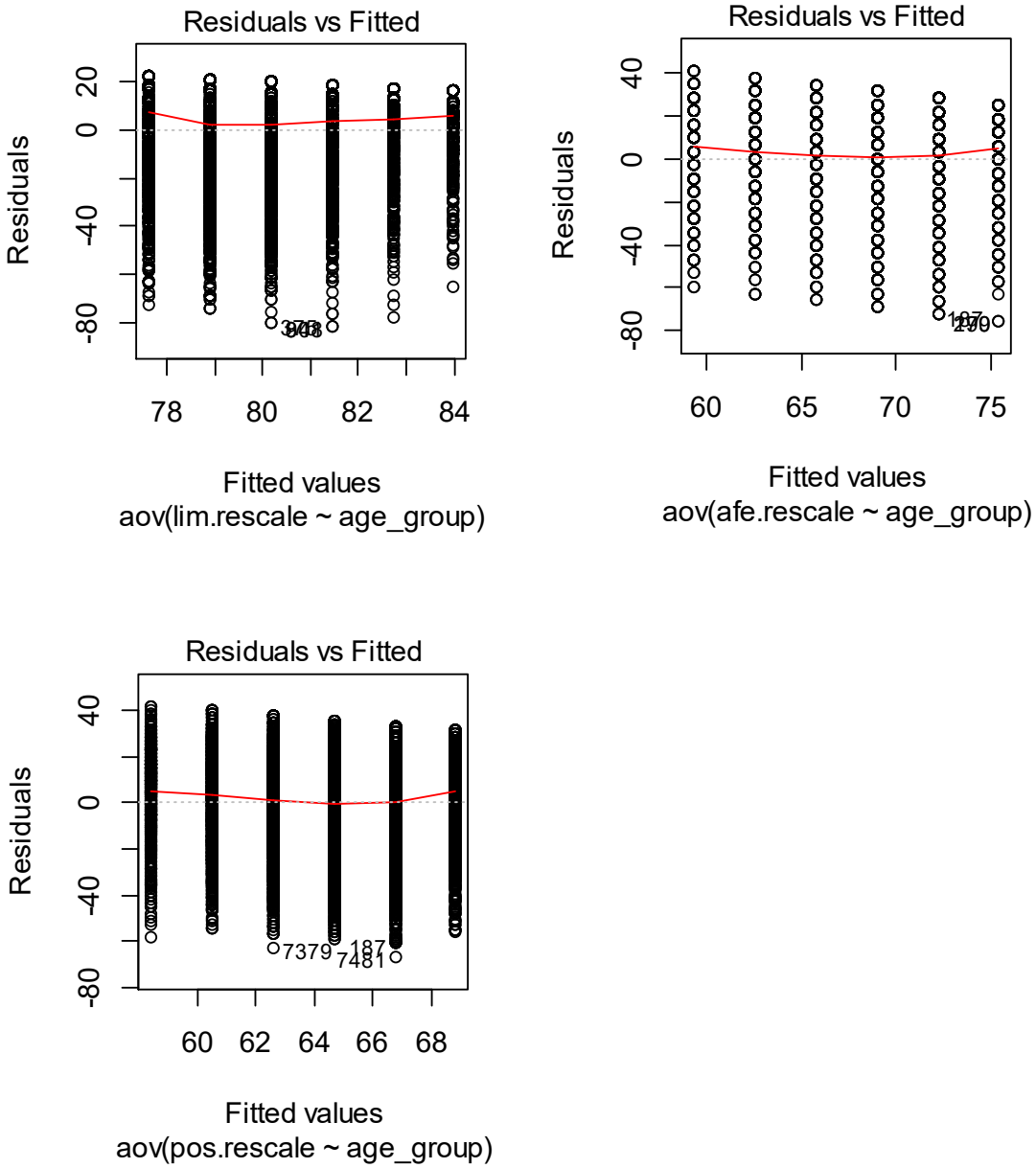
"lim.rescale" presents the values for the "Perceived Access to Options" scale, "pos.rescale" presents the values for the "Reflective Wellbeing" scale, and "afe.rescale" presents the values for the "Affective Wellbeing" scale.

Figure 5. Residuals vs. fitted values of gender subsample



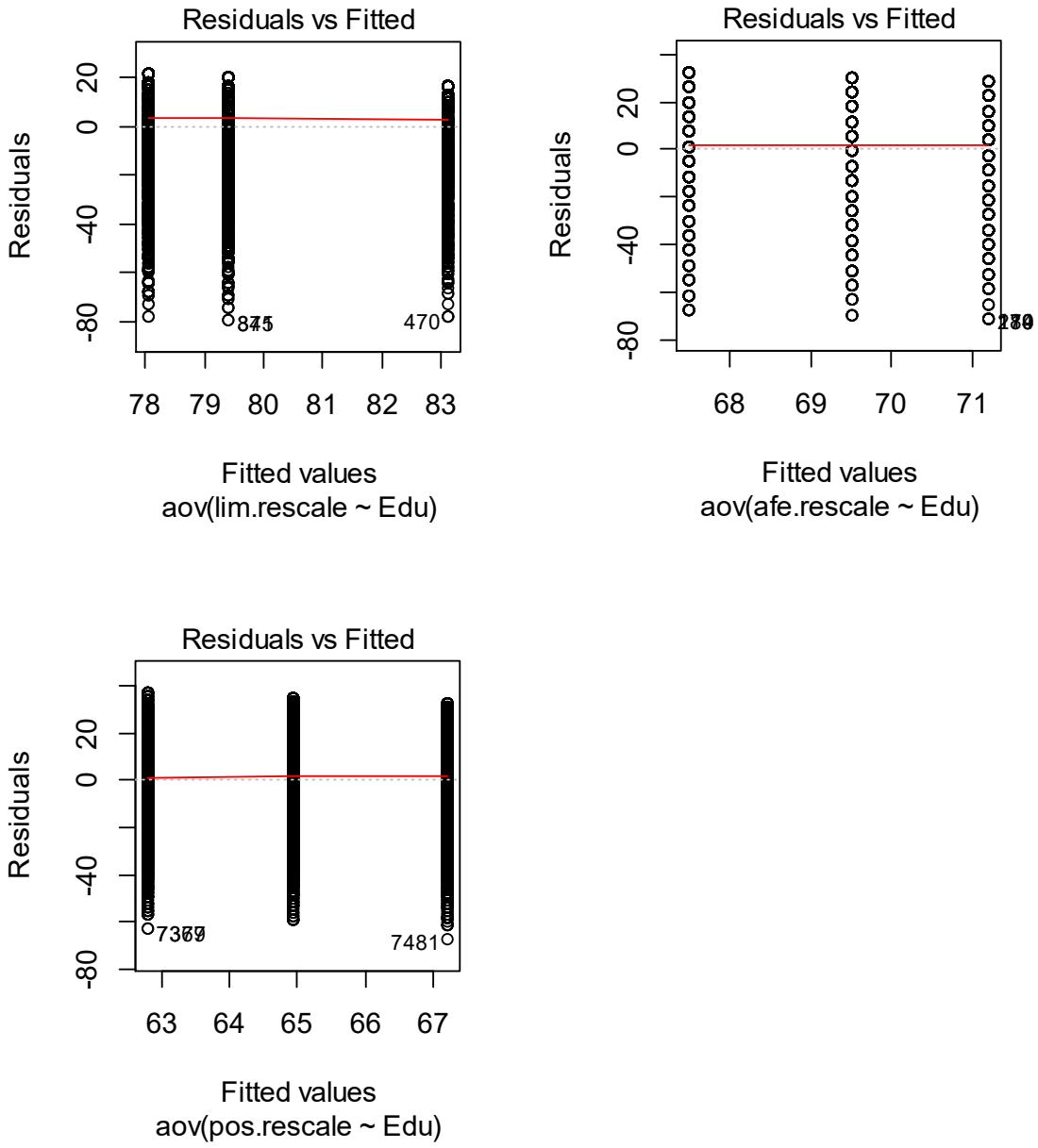
"lim.rescale" presents the values for the "Perceived Access to Options" scale, "pos.rescale" presents the values for the "Reflective Wellbeing" scale, and "afe.rescale" presents the values for the "Affective Wellbeing" scale.

Figure 6. Residuals vs. fitted values of age group subsample



"lim.rescale" presents the values for the "Perceived Access to Options" scale, "pos.rescale" presents the values for the "Reflective Wellbeing" scale, and "afe.rescale" presents the values for the "Affective Wellbeing" scale.

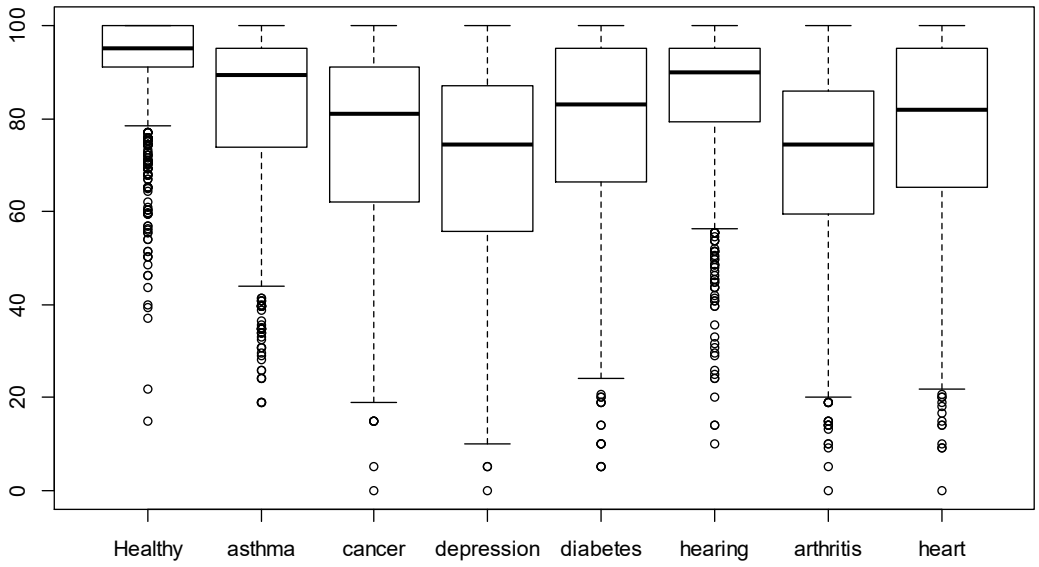
Figure 7. Residuals vs. fitted values of education subsample



"lim.rescale" presents the values for the "Perceived Access to Options" scale, "pos.rescale" presents the values for the "Reflective Wellbeing" scale, and "afe.rescale" presents the values for the "Affective Wellbeing" scale.

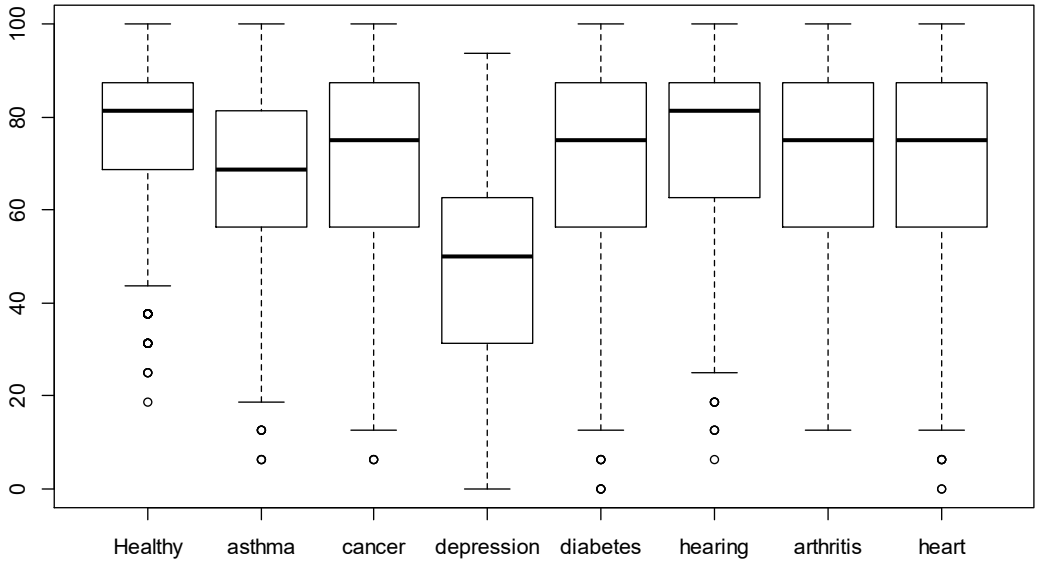
10.4.2 Boxplots for the WeRFree instrument in different subsamples.

Figure 8. Boxplot for the "Perceived Access to Options" scale for the health condition subgroups



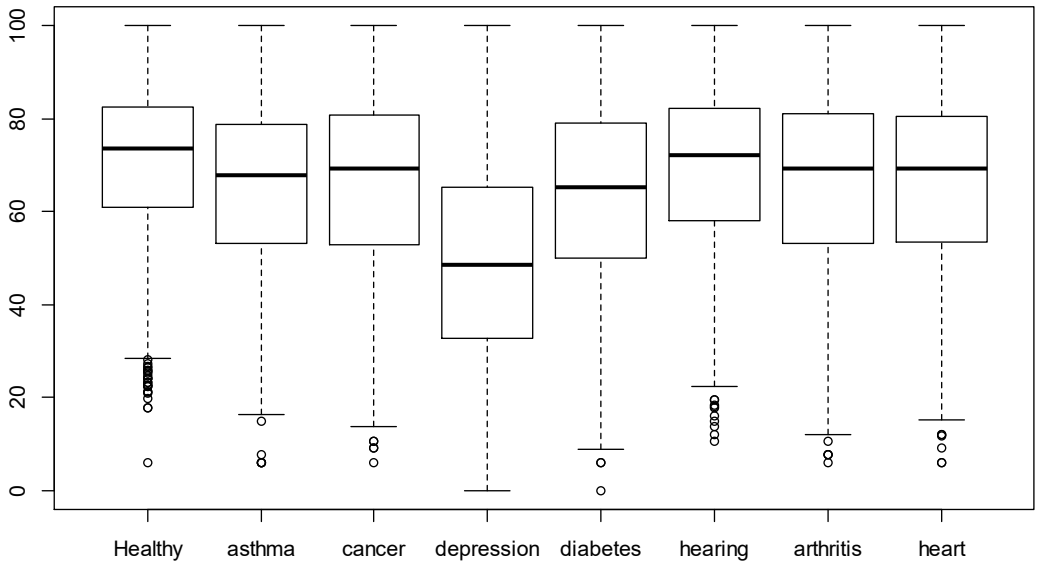
The x-axis present the different subgroups in the sample. Healthy: healthy participants, asthma: participants who have asthma, cancer: participants who have cancer, depression: participants who are depressed, hearing: participants affected by hearing loss, arthritis: participants with arthritis, heart: participants affected by heart problems. The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

Figure 9. Boxplot for the "Affective Wellbeing" scale for the health condition subgroups



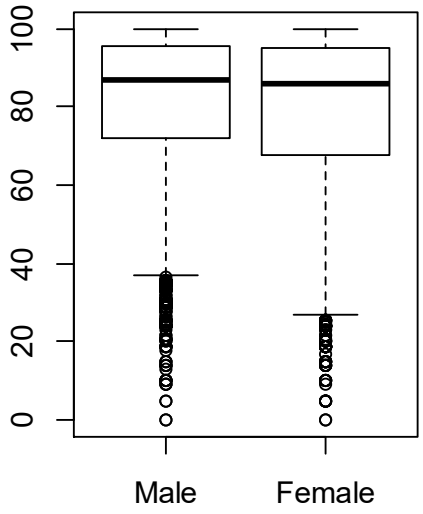
The x-axis present the different subgroups in the sample. Healthy: healthy participants, asthma: participants who have asthma, cancer: participants who have cancer, depression: participants who are depressed, hearing: participants affected by hearing loss, arthritis: participants with arthritis, heart: participants affected by heart problems. The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

Figure 10. Boxplot for the "Reflective Wellbeing" scale for the health condition subgroups



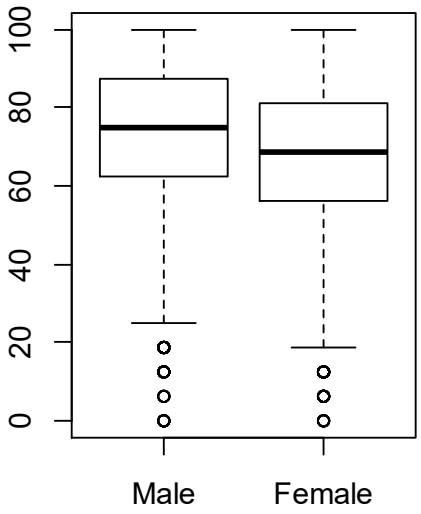
The x-axis present the different subgroups in the sample. Healthy: healthy participants, asthma: participants who have asthma, cancer: participants who have cancer, depression: participants who are depressed, hearing: participants affected by hearing loss, arthritis: participants with arthritis, heart: participants affected by heart problems. The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

Figure 12. Boxplot for the "Perceived Access to Options" scale for men and women



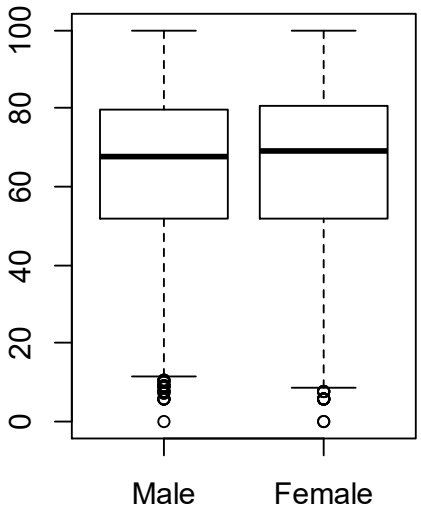
The x-axis present the different subgroups in the sample. The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

Figure 11. Boxplot for the "Affective Wellbeing“ scale for men and women



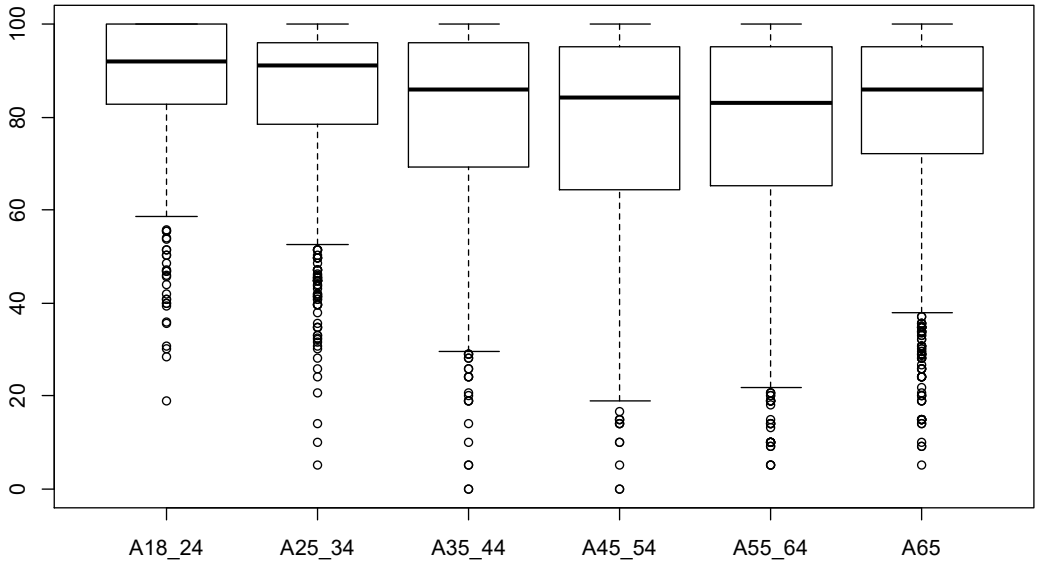
The x-axis present the different subgroups in the sample. The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

Figure 13. Boxplot for the "Reflective Wellbeing“ scale for men and women



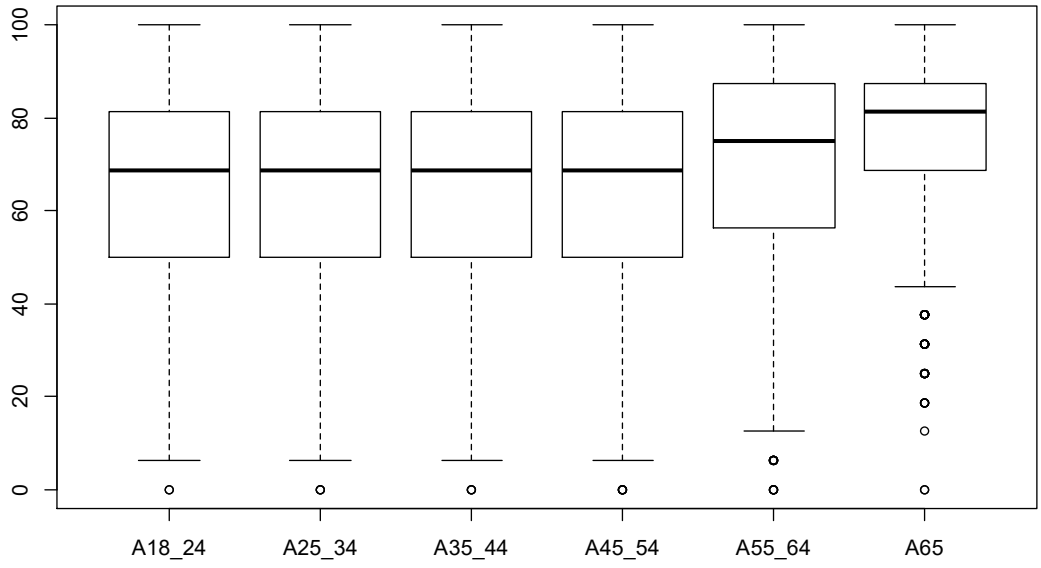
The x-axis present the different subgroups in the sample. The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

Figure 14. Boxplot for the "Perceived Access to Options“ scale for the different age subgroups



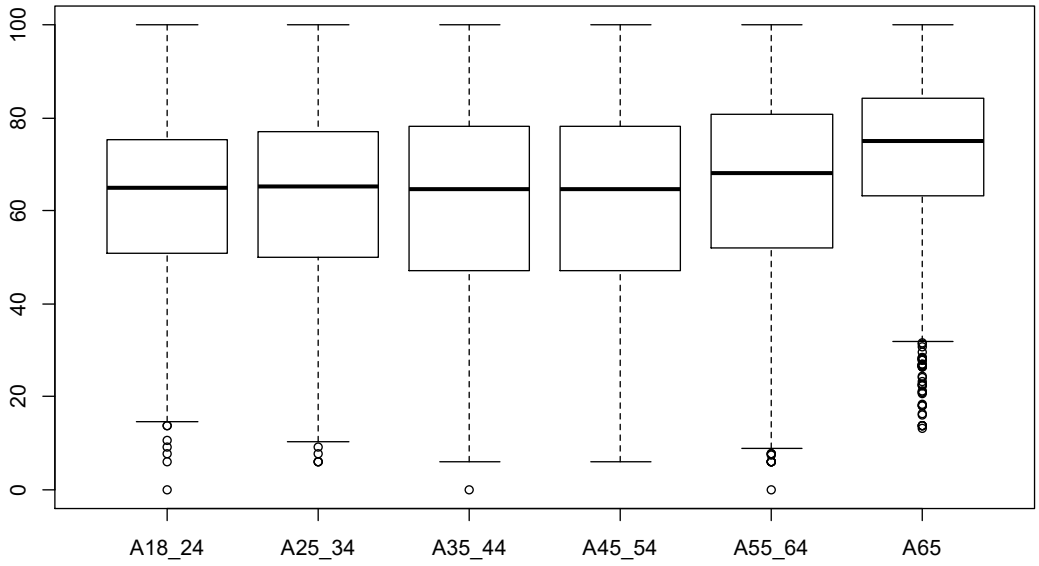
The x-axis present the different subgroups in the sample (A18_24 is participants between the age of 18 and 24, A65 is participants who are older than 65). The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

Figure 15. Boxplot for the "Affective Wellbeing“ scale for the different age subgroups



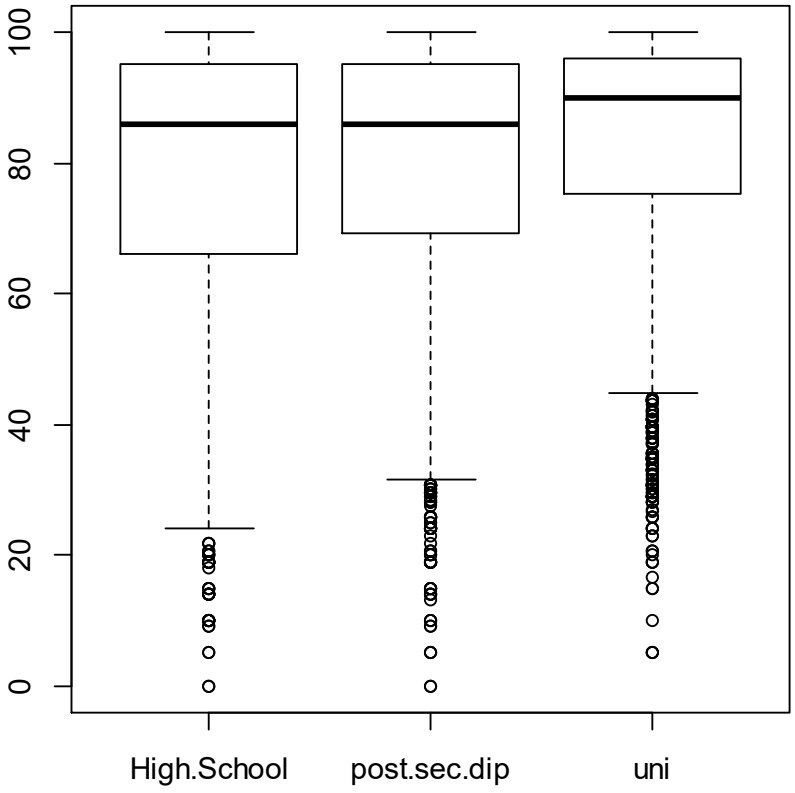
The x-axis present the different subgroups in the sample (A18_24 is participants between the age of 18 and 24, A65 is participants who are older than 65). The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

Figure 16. Boxplot for the "Reflective Wellbeing“ scale for the different age subgroups



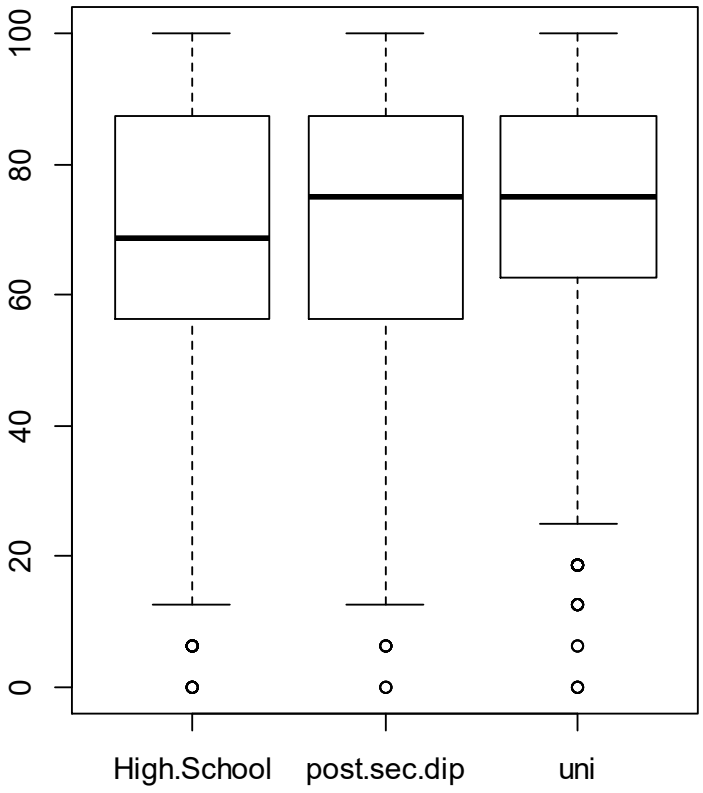
The x-axis present the different subgroups in the sample (A18_24 is participants between the age of 18 and 24, A65 is participants who are older than 65). The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

Figure 17. Boxplot for the "Perceived Access to Options“ scale for different education level subgroups



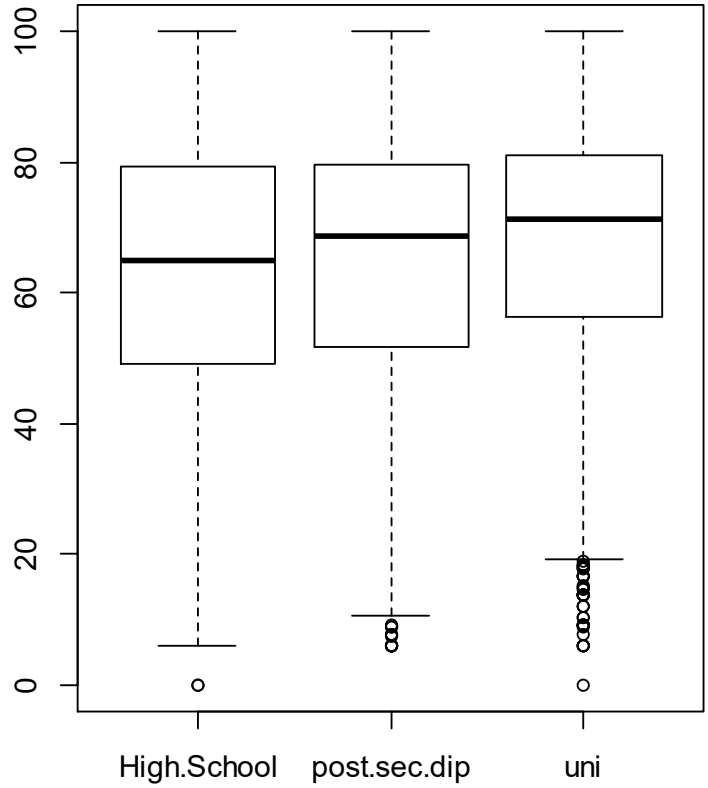
The x-axis present the different subgroups in the sample (post.sec.dip means that the participant’s highest diploma is from post-secondary education, uni means university diploma). The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

Figure 18. Boxplot for the "Affective Wellbeing“ scale for different education level subgroups



The x-axis present the different subgroups in the sample (post.sec.dip means that the participant’s highest diploma is from post-secondary education, uni means university diploma). The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

Figure 19. Boxplot for the "Reflective Wellbeing“ scale for different education level subgroups



The x-axis present the different subgroups in the sample (post.sec.dip means that the participant’s highest diploma is from post-secondary education, uni means university diploma). The y-axis presents the scale score, from 0 to 100 with 100 having optimal wellbeing.

10.5 Further validation studies of instruments

Due to a lack of time and resources, it was unfortunately not feasible to conduct extensive validation studies with the WeRFree instrument (examples can be found in step (9) in Section 1.6.1) as outlined in the guide of Boateng et al. (2018). This appendix is an illustration of how such studies could be conducted, with reference to my experience in validation studies of two questionnaires: the German version of the ICECAP-A and an instrument that was developed to measure the impact of COVID-19 on cancer patients.

10.5.1 Case study one. Back translating the ICECAP-A

10.5.1.1 Background

Sometimes instruments that have been developed in one language are applied in another, which requires a translation. A translation of an instrument into a different language is however not a straightforward exercise, due to the risk that the meaning of the constructs that the original instruments aim to measure is lost in translation. This is problematic since it becomes unclear what types of constructs the translated instrument actually measures, which hampers the interpretation of scale scores. It is therefore important that the translated version of an instrument is able to measure the same constructs as the original version of the instrument (Wild et al., 2005).

The ICECAP-A is an example of an instrument that has been translated into a multitude of languages, one of those languages being German. This translation was conducted for the MIC study. The translation was conducted by two researchers from the MIC study team, who independently from each other translated the English instrument into German after familiarizing themselves with the concepts (Linton et al., 2020). After independent translation, these members reconciled their efforts into a single translation into German with help from a third member of the MIC study. Lastly, the entire MIC study team proofread the German translation to identify any potential mistakes (Linton et al., 2020). The quality of this translation was however not formally studied. Therefore, the aim of the study presented in this section is to study the quality of the German translation of the ICECAP-A.

10.5.1.2 Methods

To study the quality of the German translation of the ICECAP-A, Linton et al. (2020) conducted an analysis in which I participated. In this analysis, two methods were used to study the quality of the translation: (1) a comparison of the psychometric performance of the English and German versions of the ICECAP-A and (2) translating the German version of the ICECAP-A back into English. I was responsible for studying the quality of the translation according to the second method. Its methods and results will be presented in this section.

One method to test whether a translation is conducted correctly is by performing a "back translation" (Wild et al., 2005). In a back translation, a newly translated version of a questionnaire is translated back into its original language. Then, the original version of the questionnaire is compared with the back-translated version (Wild et al., 2005). Differences in the wording of items between these two versions point to potential problems, such as a change in the meaning of certain items. In the study presented in this section of the Appendix, a difference was considered to be meaningful when that difference presents a fundamental difference in understanding between two items. Such differences were called "semantical". Differences were considered to be just literal (and not semantical) when they did not affect the meaning of items.

The back translation of the ICECAP-A instrument from German to English and subsequent comparison was conducted with a group of researchers working for the Division of Health Economics of the DKFZ. Two members of this group were laypeople concerning the capability approach. One of these laypeople was a native German with excellent skills in English (from here on called NatGerLay). The other was a Native English speaker who had an intermediate level of German skills (from here on called NatEngLay). The team was further supported by a third individual who is a Native German speaker with excellent skills in English and knowledge of the ICECAP-A (from here on called NatGerExp). I was the last member of this team, who fulfilled the role of project manager, due to my knowledge of the ICECAP-A, German, and English.

The back translation and its comparison to the original English version was conducted in several steps. NatGerLay was first asked to translate the German version of the ICECAP-A back into English by herself. Then, the English translation was

independently compared to the German version by NatEngLay and me to check for grammatical and semantical inconsistencies. In this comparison, NatEngLay and I listed the differences in the wording of the two versions. Differences in wording that were identified were compiled in a document that formed the basis of a topic guide for a group discussion. The group consisted of NatGerLay, NatGerExp, NatEngLay, and me. Additional topics were added that inquired NatGerLay about how understandable the German version of the ICECAP-A was.

Following the topic guide, first, the German translation of the ICECAP-A was compared with its English translation to identify any potential semantical differences. Then, the differences in the wording of items between the two English versions of the ICECAP-A were discussed to determine whether they were of semantical or literal nature. After the group discussion, a report was written and sent to the participants of the discussion to allow them to give further feedback. The largest semantical and literal differences that were identified are presented and discussed in this report.

10.5.1.3 Results

10.5.1.4 Differences in wording

Appendix Table 7 shows a comparison of the original ICECAP-A, the German version of the ICECAP-A, and the back-translated version of the ICECAP-A. The German version of the ICECAP-A was considered to be easy to understand by NatGerLay. Multiple differences in item wording were identified. The results presented in this section are a summary of the results of the group discussion.

Appendix Table 7. Comparison of the original ICECAP-A, the German translation of the ICECAP-A, and the back-translated version of the ICECAP-A

Original ICECAP-A	**German version ICECAP-A**	**Back translated version ICECAP-A**
1. Feeling settled and secure	1. Sich sicher und geborgen fühlen	1. Feeling safe and secure/loved/cared for
I am able to feel settled and secure in all areas of my life	Ich kann mich in allen Bereichen meines Lebens sicher und geborgen fühlen	I am able to feel safe and loved in all areas of my life
I am able to feel settled and secure in many areas of my life	Ich kann mich in vielen Bereichen meines Lebens sicher und geborgen fühlen	I am able to feel safe and loved in many areas of my life
I am able to feel settled and secure in a few areas of my life	Ich kann mich in nur in wenigen Bereichen meines Lebens sicher und geborgen fühlen	I am able to feel safe and loved in only few areas of my life
I am unable to feel settled and secure in any areas of my life	Ich kann mich in keinem Bereich meines Lebens sicher und geborgen fühlen	I am not able to feel safe or loved in any area of my life
2. Love, friendship and support	2. Liebe, Freundschaft und Rückhalt	2. Love, friendship and support
I can have a lot of love, friendship and support	Ich kann auf viel Liebe, Freundschaft und Rückhalt zurückgreifen	I am able to rely on a lot of love, friendship and support
I can have quite a lot of love, friendship and support	Ich kann auf ziemlich viel Liebe, Freundschaft und Rückhalt zurückgreifen	I can rely on quite a lot of love, friendship and support
I can have a little love, friendship and support	Ich kann auf nur wenig Liebe, Freundschaft und Rückhalt zurückgreifen	I can rely on a little love, friendship and support
I cannot have any love, friendship and support	Ich kann auf gar keine Liebe, Freundschaft und Rückhalt zurückgreifen	I cannot rely on love, friendship or support at all

3. Being independent I am able to be completely independent I am able to be independent in many things I am able to be independent in a few things I am unable to be at all independent	3. Unabhängig sein Ich kann vollständig unabhängig sein Ich kann in vielerlei Hinsicht unabhängig sein Ich kann nur in gewissen Dingen unabhängig sein Ich kann nicht unabhängig sein	3. Being independent I can be fully independent I can be independent in many ways I can only be independent in certain matters I cannot be independent.
4. Achievement and progress I can achieve and progress in all aspects of my life I can achieve and progress in many aspects of my life I can achieve and progress in a few aspects of my life I cannot achieve and progress in any aspects of my life	4. Leistungen und Vorankommen Ich kann in allen Bereichen meines Lebens etwas leisten und vorankommen Ich kann in vielen Bereichen meines Lebens etwas leisten und vorankommen Ich kann nur in wenigen Bereichen meines Lebens etwas leisten und vorankommen Ich kann in keinem Bereich meines Lebens etwas leisten und vorankommen	4.AchievementspPerformance/accomplishments and moving forward/moving ahead/progress I can achieve things and make progress in all areas of my life I can achieve things and make progress in many areas of my life I can achieve things and make progress in few areas of my life I cannot achieve nor make progress in any area of my life
5. Enjoyment and pleasure I can have a lot of enjoyment and pleasure I can have quite a lot of enjoyment and pleasure I can have a little enjoyment and pleasure I cannot have any enjoyment and pleasure	5. Vergnügen und Genuss Ich kann viel Vergnügen und Genuss erleben Ich kann ziemlich viel Vergnügen und Genuss erleben Ich kann nur wenig Vergnügen und Genuss erleben Ich kann kein Vergnügen und Genuss erleben	5. Fun/amusement/joy/pleasure and enjoyment/pleasure (food, body & mind) I am able to experience a lot of amusement/fun and enjoyment I am able to experience quite a lot of amusement/fun and enjoyment I am able to experience just a little bit of amusement/fun and enjoyment I am not able to experience any amusement/fun or enjoyment

No large semantical differences between the German and English versions of the ICECAP-A were identified in the group discussion. Indeed, the items of the two versions generally reflected the same concepts. Smaller differences in the semantics of items were however identified in items 1 and 5. In item 1, both NatGerLay and NatGerExp remarked that the word “geborgen” is more strongly related to attachment and an interpersonal sense of stability, while in the original ICECAP-A this question represented a more individual sense of stability (lay terms associated with this item according to the ICECAP-A development paper (Al-Janabi et al., 2012): stable, settled, secure, not worried, relaxed, comfortable). There is some overlap between item 1 and item 2 due to this social component in question 1 in the back translation, which is not represented in the original version of the ICECAP-A. However, the word “Sicher” in the same question captures a part of the meaning of the original question well, therefore reducing the potential that a slight difference in understanding is introduced by the word “geborgen”.

In item 5, the word “Genuss” is more related to everyday bodily enjoyments (such as food), while seeming to miss the broader range of enjoyment that the original ICECAP-A tries to capture (ranging from quiet pleasures to things in life that are fun or exciting). However, the word “Vergnügen”, seems to capture this broader range well, which indicates that the German item still manages to capture the original meaning of the question.

Besides the comparison of the German and the English versions of the ICECAP-A, also several differences in wording were identified between the English version of the ICECAP-A and its back-translated version. Items from the original ICECAP-A with the wording “I can” and “I am able to” were translated into “ich kann” in the German version. The wording “ich kann” was back-translated as “I can” and “I am able to”. However, doing so resulted in differences between the original ICECAP-A and the back-translation in items 1, 3, and 5. For these items, the original “I can” was for example translated to “I am able to” in the back-translated version. Similarly, items that originally were words such as “I am able to” were back-translated as “I can”. With respect to these differences, the members of the group considered that they were not semantically meaningful. Based on this, the members argued that the German translation “ich kann...” sufficiently reflects the meaning of “I can...” and “I am able to...”.

In item 2, the original ICECAP-A asks, “I can…”. This was back-translated as “I can rely on”. This is a result from the German translation “Ich kann auf … zurückgreifen”. Literally, there was a difference between the wordings of the two items in the back translation. However, both NatGerLay and NatGerExp agreed that a semantical difference does not exist between the German version of the ICECAP-A and the English version of the ICECAP-A, given that the meaning of the original English item is correctly conveyed. The members of the group discussion agreed that the German version reflects the meaning of the original item 2 sufficiently well.

In item 4, The English back translation uses the wording “I can achieve things ...”. In the original English version, however, item 4 uses the wording “I can achieve ...”. During the group discussion, it was clarified that the word “things” was added by NatGerLay as a result of how German sentences are grammatically structured. Also, it was noted that the German version of the ICECAP-A does not mention a phrase similar to the word “things”. The group decided the translation of the ICECAP-A is fine as it is.

10.5.1.5 Discussion

Overall, the comparison of the German translation and the back translation of the ICECAP-A to the original English version did not indicate that there are fundamental semantic differences. Some smaller semantic differences between the original ICECAP-A and the back translation were identified in items 1 and 5, however, the members of the group discussion did not consider these differences to result in a fundamentally different interpretation of the items.

One limitation of back translation is that no professional translators were asked to conduct the back translation due to limited resources. The use of professional translators is the golden standard for this type of study (Wild et al., 2005). We aimed to minimize this limitation by conducting a group discussion that consisted of a mix of laypeople and experts in terms of the ICECAP-A and the English and German languages. Still, it would be interesting to study whether a professional translator would come with a fundamentally different back translation.

Besides the back translation, also further psychometric analysis was part of the study by Linton et al. (2020) in which I participated. The English and German versions of the

ICECAP-A showed that both versions are similar in terms of reliability (Cronbach's alpha German version = 0.83, Cronbach's alpha English version = 0.85). Furthermore, also the convergent validity of the two versions of the ICECAP-A was similar in terms of the correlations of their scores to the SWLS, EQ-5D-5L, and items from the SF-36 V2. The conclusion of the group discussion that German and English versions of the ICECAP-A are semantically similar is thus supported by further psychometric analysis.

10.5.2 Case study two. Pre-testing a socioeconomic impact questionnaire

10.5.2.1 Introduction

The coronavirus disease 2019 (COVID-19) pandemic has had an unprecedented impact on the lives of people. A group that might particularly be affected are cancer patients. The Krebsinformationsdienst (KID), an organization that people can contact for any type of cancer-related question, reported an increase in questions from cancer patients about a range of issues since the start of the pandemic, such as access to care, stress, and uncertainty, as well as financial hardship. However, at the time there was no study on the frequency of changes in healthcare services and the psychological impact of such changes.

For this reason, researchers from the DKFZ (from the Division of Health Economics and the KID), and the National Center for Tumor Diseases Heidelberg initiated the "Impact of the Corona Crisis on Cancer Patients: medical, psychosocial and economic consequences and their ethical implications" (CoCan-ELSI) project. The aim of the CoCan-ELSI project is to study the effects of the corona-crisis on healthcare services for cancer patients, the QoL of cancer patients, and the financial situation of cancer patients. To study these effects, a questionnaire was developed by researchers from the DKFZ and the National Center for Tumor Diseases Heidelberg (NCT). This questionnaire is a mix of validated instruments and items that are self-developed by the research term. However, it was unclear whether the self-developed items are of sufficient quality (i.e., if they are understandable for participants and if they measure what they intend to measure). The aim of the study presented in this section is therefore to conduct a pre-test with these items.

10.5.2.2 Methods

Cognitive pre-tests were conducted to study if the self-developed questions are understandable and measure what they are intended to measure (the used questionnaire to test the self-developed questions can be found in Section 10.5.4.1). These pre-tests were conducted according to the concurrent think-aloud methodology (Lenzner et al., 2016). Participants were asked and encouraged to express their thoughts out loud while they were answering the self-developed questions. At times, interviewers asked probing questions to ensure that they followed the thoughts of the participants. Meanwhile, the interviewers investigated if the participants had any problems responding to items with respect to the following questions (Lenzner et al., 2016):

1. Are the questions understandable for the participants?
2. How do participants retrieve the necessary information to answer the question from their memory?
3. How do participants evaluate the retrieved information to come to a response?
4. Do the participants correctly match their internal responses with the answers provided in the questionnaire?

If responses to the items showed problems in any of these questions, notes were taken that described what the problem was with the respective items. When applicable, feedback was requested from the participants about how they would improve an item. After the pre-tests, the interviewers would review the notes and the feedback of the participants to come up with suggestions to improve the questionnaire.

The questionnaire for the CoCan-ELSI project was developed with a team of 10 members from the KID, the Division of Health Economics of the DKFZ, and the NCT. The questionnaire consists of validated instruments and self-developed items. The pre-tests concerned these self-developed items (the original questionnaire used for the pre-test can be found in Section 10.5.4.1). These pre-tests were conducted by two members of the DKFZ: AG and me. AG is a physician working for the KID. She has conducted surveys on several topics. Amongst other things, these topics cover cancer patients´ information and healthcare communication needs and perceptions, health literacy, and shared decision-making. Through this research, AG has experience in the

development and pre-testing of instruments. I was a doctoral student in the Division of Health Economics. Through my doctoral research, I also have experience in the quantitative and qualitative testing and validation of instruments.

The pre-tests started with an introduction, in which the interviewers explained the reasons for conducting the pre-test, the aim of the questionnaire, and the structure of the questionnaire. Further explanation was given about what was expected from the participants in the think-aloud method. Additionally, it was stressed that no recordings were made and that the researchers would only take notes about the quality of the items. Participants were also informed that the pre-tests were anonymous and they could quit at any time during the pre-test. Items from validated instruments that did not require pre-testing were used to help participants get used to thinking out loud. When participants remained quiet for longer than a couple of seconds, the interviewers asked the participants to share their thoughts.

Each interviewer individually interviewed five participants. After the interviews were finished, the interviewers discussed the resulting feedback and notes, which resulted in a summary report about the quality of individual items with suggestions for improvement. This summary was presented and discussed with the full team of the CoCan-ELSI project. The aim of this discussion was to assess how the instrument can be improved. When necessary, items were reformulated or restructured, extra information was provided, and answer categories were added, reformulated, or removed. This resulted in a final version of the instrument.

Because of the urgency dictated by the corona crisis, a convenience sample of ten participants was recruited from the personal networks of the interviewers. In the recruitment process, AG and JU aimed to recruit participants with different backgrounds. The questionnaire was developed and pre-tested in April and June 2020.

10.5.2.3 Results

The complete sample consisted of five women and five men. The participants interviewed by me were in an age range from 21 to 34. The participants interviewed by AG were in an age range from 60 to 75. The ten participants came from a range of different occupational backgrounds (journalism, teaching, nursing, psychology, media design, engineering, and health-related studies). At the time of conducting the

interviews, the participants had not been infected by COVID-19, positively tested for COVID-19, nor were otherwise symptomatic. They were also not affected by cancer or its treatment. However, some participants mentioned that they knew people who were affected by cancer through their work (for example a participant who worked as a nurse) or knew people in their social circle. The pre-tests took 45 to 90 minutes to conduct.

The participants reported that the self-developed items were generally intuitive and well-structured. Furthermore, the participants observed that the self-developed items considered most of the important aspects related to the impact of COVID-19 on the care of patients, the QoL of patients, and the financial situation of patients. Still, some of the items could be improved, for which notes were made.

The notes associated with these items were evaluated by the developers of the instrument from the DKFZ and the NCT, as well as by AG and me, which resulted in improvements to these items. These improvements were considered minor, given that they did not fundamentally change the meaning of individual items in the questionnaire; rather, changes to items mostly concerned changing or adding additional response options to items, as well as improving the wording of items such that the items more closely reflect what the researchers intended to measure. Examples of such include the addition of a “Wohngemeinschaft” (a shared apartment) option to an item that asks about the type of housing in which participants live, or further clarifications of the meaning of various clinical terms, such as what is meant with “Nachsorge” (care after an intervention).

Not all the notes were used to change items, since the researchers believed that some difficulties in responding to the items were related to participants themselves not being affected by cancer and its treatment. The improved version of the questionnaire can be found in Section 0. Its format and the order of the questions have been slightly changed to make the questionnaire more feasible to use online. Additional instructions can also be found in italic, which further clarify how participants were guided through the questions.

10.5.2.4 Discussion

Some of the self-developed items had minor issues that could be resolved. Following the results of the pre-test and the subsequent improvement of the questionnaire, the researchers were confident that the survey is ready to be implemented from a technical perspective.

Still, it should be noted that the target group of this instrument are people who are affected by cancer. The participants that were part of the pre-tests were not affected by cancer themselves. At times, this resulted in difficulties for the participants to understand what the items exactly inquired, since the participants did not have experience with the various treatment options, the impact of cancer on their own QoL, as well as the impact of cancer on their own financial situation. It is therefore possible that certain items might not completely reflect the experiences of cancer patients.

The instrument was used in a research project, in which patients who contacted the KID were asked to complete the survey (Eckford et al., 2021; Gaisser et al., 2022). In total 621 patients participated, of whom 13% reported a change in their treatment or care (Eckford et al., 2021). The patients who experienced these changes reported higher levels of anxiety and depression (Eckford et al., 2021). In light of these findings, one of the study's conclusions was that healthcare professionals and policymakers should pay additional attention to the psychological and social wellbeing of patients (Eckford et al., 2021).

10.5.3 Overall discussion

When reflecting on the guidelines by Boateng et al. (2018), it should be noted that there is still much work left before the WeRFree instrument can be used to assess wellbeing. Regarding the WeRFree instrument, the back translation presented in section 10.5.1 might be slightly less relevant, since the items of the instrument are based on validated studies that have already been formally translated into several languages.

However, a cognitive think-aloud study, as presented in Section 10.5.2, might be interesting to study whether the responses of individuals reflect the constructs of the WeRFree instrument as intended. This is particularly interesting in the case of the

“Perceived Access to Options” scale. In this dissertation, I have tried to show that at least on a theoretical level there is a relation between the concept of option freedom and various items of HRQoL instruments. A cognitive think-aloud study might provide further evidence of whether individuals that respond to these items evaluate their health-related option freedom. Such qualitative studies could thus provide an insight into the content and construct validity of the WeRFree instrument.

10.5.4 Questionnaire used for pre-test

10.5.4.1 The original questionnaire used for the pre-test.

Befragung zu den Folgen der Corona-Pandemie für Krebspatienten

Corona-Pandemie und Gesundheitsversorgung

Jetzt interessiert uns, ob die Corona-Pandemie Auswirkungen auf Ihre medizinische Versorgung hatte oder hat.

Sind Sie bzw. waren Sie selbst an COVID-19 erkrankt?

- o Nein, ich bin weder aktuell noch war ich an Covid-19 erkrankt.
- o Ja, ich bin aktuell an COVID-19 erkrankt.
- o Ja, ich war an COVID-19 erkrankt und bin inzwischen wieder genesen.
- o Mir ist nicht bekannt, ob ich mit COVID-19 infiziert bin oder war.

Ist oder war jemand in Ihrer Familie oder Ihrem Freundeskreis an COVID-19 erkrankt?

- o Ja, in meiner Familie
- o Ja, in meinem engen Freundeskreis
- o Nein bzw. weiß nicht

Hat sich durch die Corona-Pandemie etwas in Ihrem Behandlungs- oder Nachsorgeplan geändert?

- o Nein, es hat sich nichts geändert
- o Ja, es hat sich etwas geändert
  (*Bitte kreuzen Sie an, bei welcher Maßnahme/bei welchen Maßnahmen eine Änderung erfolgt ist (Mehrfachnennungen möglich).*

- O Operation
- O Medikamentöse Tumortherapie (z. B. Chemotherapie, Hormontherapie, zielgerichtete Therapie, Immuntherapie)
- O Bestrahlung
- O Verlaufskontrolle während der Behandlung
- O Nachsorgetermin
- O Psychosoziale / psychoonkologische Betreuung
- O Pflegerische Betreuung

- O **Operation**

  - o Verschoben
  - o Entfällt
  - o Modifiziert
    Art der Modifikation: _____

Was war der Grund?
- o Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
- o Versorgungsengpass wegen der Corona-Pandemie
- o Mein Gesundheitszustand
- o Die Operation ist nicht dringend
- o Anderer Grund: _____
- o Weiß nicht

Für wie lange wurde verschoben?
- o Bis 4 Wochen
- o Länger als 4 Wochen
- o Auf unbestimmte Zeit
- o Entfällt ganz

Wer hat die Entscheidung getroffen?
- o Ich selbst
- o Die behandelnden Ärzte
- o Die behandelnden Ärzte und ich gemeinsam

Wie sehr belastet Sie diese Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

O Medikamentöse Tumortherapie (z. B. Chemotherapie, Hormontherapie, zielgerichtete Therapie, Immuntherapie)

- o Verschoben
- o Entfällt

Was war der Grund?
- o Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
- o Versorgungsengpass wegen der Corona-Pandemie
- o Mein Gesundheitszustand
- o Die Therapie ist nicht dringend
- o Anderer Grund: _____
- o Weiß nicht

Für wie lange wurde verschoben?
- o Bis 4 Wochen
- o Länger als 4 Wochen
- o Auf unbestimmte Zeit
- o Entfällt ganz

Wer hat die Entscheidung getroffen?
- o Ich selbst
- o Die behandelnden Ärzte
- o Die behandelnden Ärzte und ich gemeinsam

Wie sehr belastet Sie diese Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

o **Bestrahlung**

o Verschoben
o Entfällt

Was war der Grund?
o Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
o Versorgungsengpass wegen der Corona-Pandemie
o Mein Gesundheitszustand
o Die Bestrahlung ist nicht dringend
o Anderer Grund: _____
o Weiß nicht

Für wie lange wurde verschoben?
o Bis 4 Wochen
o Länger als 4 Wochen
o Auf unbestimmte Zeit
o Entfällt ganz

Wer hat die Entscheidung getroffen?
o Ich selbst
o Die behandelnden Ärzte
o Die behandelnden Ärzte und ich gemeinsam

Wie sehr belastet Sie diese Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

o **Verlaufskontrolle während der Behandlung**

o Verschoben
o Entfällt
o Umgestellt auf Telefongespräch

Was war der Grund?
o Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
o Versorgungsengpass wegen der Corona-Pandemie
o Mein Gesundheitszustand
o Die Untersuchungen sind nicht dringend
o Anderer Grund: _____
o Weiß nicht

Für wie lange wurde verschoben?
o Bis 4 Wochen
o Länger als 4 Wochen
o Auf unbestimmte Zeit
o Entfällt ganz

Wer hat die Entscheidung getroffen?
o Ich selbst
o Die behandelnden Ärzte
o Die behandelnden Ärzte und ich gemeinsam

Wie sehr belastet Sie diese Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

o **Nachsorgetermin**

o Verschoben
o Entfällt
o Umgestellt auf Telefongespräch

Was war der Grund?
o Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
o Versorgungsengpass wegen der Corona-Pandemie
o Mein Gesundheitszustand
o Der Nachsorgetermin ist nicht dringend
o Anderer Grund: _____
o Weiß nicht

Für wie lange wurde verschoben?
o Bis 4 Wochen
o Länger als 4 Wochen
o Auf unbestimmte Zeit
o Entfällt ganz

Wer hat die Entscheidung getroffen?
o Ich selbst
o Die behandelnden Ärzte
o Die behandelnden Ärzte und ich gemeinsam

Wie sehr belastet Sie die Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

O **Psychosoziale / psychoonkologische Betreuung**

- O Verschoben
- O Entfällt
- O Umgestellt auf Telefongespräch

Was war der Grund?

- O Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
- O Versorgungsengpass wegen der Corona-Pandemie
- O Mein Gesundheitszustand
- O Die Betreuung ist nicht dringend
- O Anderer Grund: _____
- O Weiß nicht

Für wie lange wurde verschoben?

- O Bis 4 Wochen
- O Länger als 4 Wochen
- O Auf unbestimmte Zeit
- O Entfällt ganz

Wer hat die Entscheidung getroffen?

- O Ich selbst
- O Mein Betreuer/ Therapeut
- O Mein Betreuer/ Therapeut und ich gemeinsam

Wie sehr belastet Sie die Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

O **Pflegerische Betreuung**
- O Verschoben
- O Entfällt

Was war der Grund?
- O Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
- O Versorgungsengpass wegen der Corona-Pandemie
- O Mein Gesundheitszustand
- O Die Betreuung ist nicht dringend
- O Anderer Grund: _____
- O Weiß nicht

Für wie lange wurde verschoben?
- O Bis 4 Wochen
- O Länger als 4 Wochen
- O Auf unbestimmte Zeit
- O Entfällt ganz

Wer hat die Entscheidung getroffen?
- O Ich selbst
- O Mein Betreuer/ Therapeut
- O Mein Betreuer/ Therapeut und ich gemeinsam

Wie sehr belastet Sie die Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

Wurde Ihnen wegen der Änderungen Unterstützung / Beratung angeboten?
- O Nein
- O Ja

Durch wen? *(Bitte ankreuzen; Mehrfachnennungen möglich)*
- O Durch meine Krankenversicherung
- O Durch meinen behandelnden Arzt
- O Durch einen Psycho-Onkologen
- O Durch einen anderen Arzt / eine andere Klinik
- O Durch eine Beratungsstelle / Beratungsdienst
- O Anderes (bitte angeben): _____

Wie beurteilen Sie selbst die Veränderungen?

	Stimme voll und ganz zu	Stimme zu	Weder noch	Stimme nicht zu	Stimme gar nicht zu
1. Angemessen	O	O	O	O	O
2. Beeinträchtigen (möglicherweise) den Behandlungserfolg	O	O	O	O	O

III Lebensqualität und Belastung

	Immer	Meistens	Manchmal	Selten	Nie
Wie oft haben Ihre körperliche Gesundheit oder Ihre seelischen Probleme in den letzten 4 Wochen Ihre Kontakte zu anderen Menschen beeinträchtigt? *(Gemeint sind nicht allgemeine Kontaktbeschränkungen wegen der Corona-Krise)*.	O	O	O	O	O

Jetzt geht es darum, welche Veränderungen sich durch die Corona-Pandemie möglicherweise auf ihr Befinden ergeben haben.
Bitte bewerten Sie für den Zeitraum der letzten 4 Wochen, inwieweit die folgenden Aussagen für Sie zutreffen.

Als Folge der Coronavirus-Krise, mache ich mir Sorgen...	**trifft ganz genau zu**	**trifft eher zu**	**teils-teils**	**trifft eher nicht zu**	**trifft gar nicht zu**
a. ... wegen finanzieller Belastungen durch meine Krebserkrankung.	O	O	O	O	O
b. ... um meine Arbeitssituation und die finanziellen Folgen.	O	O	O	O	O
c. ... dass meine Finanzsorgen nicht nur kurzdauernd sind.	O	O	O	O	O
d. ... selbst an Covid19 zu erkranken.	O	O	O	O	O
e. ... wegen möglicher Auswirkungen auf die Qualität der medizinischen Versorgung meiner Krebserkrankung.	O	O	O	O	O

Wie erleben Sie die Einschränkungen persönlicher Kontakte und Begegnungen durch die Corona-Pandemie? *(Bitte bewerten Sie, wie sehr Ihnen die verschiedenen Kontakte fehlen.)*

	fehlen mir sehr	**fehlen mir eher**	**teils-teils**	**fehlen mir eher nicht**	**fehlen mir gar nicht**
a. Kontakte mit Angehörigen, Nachbarn und Freunden …	O	O	O	O	O

b. Kontakte mit anderen Krebspatienten/mit meiner Selbsthilfegruppe …	O	O	O	O	O
c. Kontakte mit Ärzten und Pflegenden …	O	O	O	O	O
d. Kontakte mit Betreuern, Therapeuten und anderen Helfern …	O	O	O	O	O
e. Kontakte in der Öffentlichkeit …	O	O	O	O	O

16. Was hilft Ihnen mit der aktuellen Situation umzugehen? *(Mehrfachnennungen möglich)*

O Ich habe keine Probleme mit der Situation

o Ich nutze professionelle Unterstützung.
o Ich telefoniere/chatte/benutze Sozial Medien mit Familie, Freunden und Bekannten.
o Ich finde Frieden in der Natur/Gartenarbeit.
o Mein Haustier/meine Haustiere bringen mir Trost und Freude.
o Ich finde Trost in meiner Religion.
o Mein Optimismus und meine Zuversicht geben mir Kraft.
o Ich betreibe Sport, Entspannungsübungen oder Meditation.
o Anderes (bitte angeben): _____

<u>IV Finanzielle Folgen der Krebserkrankung und der Corona-Pandemie</u>

Neben den körperlichen und seelischen Belastungen kann eine Krebserkrankung auch Auswirkungen auf die finanzielle Situation haben. Die Corona-Pandemie könnte zusätzliche Kosten oder Einbußen verursachen. Wir möchten erfahren, wie das bei Ihnen ist.

17. Wie ist Ihr derzeitiger Erwerbsstatus?

o Angestellt beschäftigt
o Selbständig oder freiberuflich tätig
o Beamter/Beamtin
o Teilzeiterwerbstätig
o Berentet
o Nicht erwerbstätig
o Möchte ich nicht beantworten

Waren Sie <u>aufgrund Ihrer Krebserkrankung</u> (ohne Berücksichtigung eventueller Folgen der Corona-Krise) in den letzten 4 Wochen arbeitsunfähig?

o Nein
o Ja

Wie lange? *(bitte der Gesamtzahl der bisherigen AU-Tage in der letzten 4 Wochen hier angeben*)

_____

Haben/hatten Sie infolge Ihrer Krebserkrankung (ohne Berücksichtigung eventueller Folgen der Corona-Krise) zusätzliche Ausgaben im Zusammenhang mit Ihrer Gesundheitsversorgung?

o Nein
o Ja, nämlich *(Mehrfachnennungen möglich)*
  o Zuzahlungen für ambulante oder stationäre Leistungen
  o Fahrtkosten
  o Inanspruchnahme von Haushaltshilfen
  o Behandlungen und Medikamente, die nicht von der Krankenkasse erstattet werden
  o Pflegekraft/Pflegedienst
  o Sonstiges:_____

o Weiß nicht
o Dazu möchte ich keine Angabe machen

Wie hoch sind diese Mehrausgaben etwa monatlich?

  o Unter 100 Euro
  o 100 – 200 Euro
  o 201 – 500 Euro
  o 501 – 800 Euro
  o 801 – 1.200 Euro
  o Über 1.200 Euro
  o Dazu möchte ich keine Angabe machen

Hat die Krebserkrankung dazu geführt, dass Sie im Alltag mehr sparen müssen?

O Nein
o Ja
o Dazu möchte ich keine Angaben machen

Wo versuchen Sie Geld einzusparen? *(Mehrfachnennungen möglich)*
O In meiner Freizeitgestaltung (z. B. Kinobesuche, Essen-/Ausgehen, Reisen, Sport)
O Bei Frisör, Pflege, Kosmetik
O Bei meiner Ernährung/bei Lebensmitteln
O Bei medizinischen Behandlungen/Zusatzleistungen
O Beim Kauf von Bekleidung
O Bei Genussmitteln
O Bei Ausstattung für Haushalt und Wohnen
O Sonstiges: _____

Mögliche zusätzliche Auswirkungen der Corona-Pandemie

Sind Ihnen in den letzten 4 Wochen durch die Coronavirus-Pandemie zusätzliche Kosten oder Zeitaufwand für Ihre Gesundheitsversorgung entstanden?

- o Nein
- o Ja
- o Weiß nicht
- o Dazu möchte ich keine Angabe machen

Haben Sie Leistungen auf eigene Kosten in Anspruch genommen?

Ja, nämlich:
- o Private medizinische Behandlung
- o Komplementäre oder alternative Behandlungsmethoden
- o Labordiagnostik/apparative Untersuchungen
- o Andere (Bitte angeben): _____

Wie viel Geld haben Sie für diese Leistungen in den letzten 4 Wochen etwa ausgegeben?

- o Unter 100 Euro
- o 100 – 200 Euro
- o 201 – 500 Euro
- o 501 – 800 Euro
- o 801 – 1.200 Euro
- o Über 1.200 Euro
- o Dazu möchte ich keine Angabe

Sind Ihnen durch die Corona-Pandemie Einkommenseinbußen entstanden, z.B. durch die Aufgabe eines (Neben-)Jobs, Verringerung der Arbeitszeit, etc.?

- o Nein
- o Ja, monatlich in Höhe von
  - o Unter 100 Euro
  - o 100-200 Euro
  - o 201-500 Euro
  - o 501-800 Euro
  - o 801-1.200 Euro
  - o Über 1.200 Euro
- O Weiß nicht
- O Dazu möchte ich keine Angabe machen

Haben/hatten Sie die Möglichkeit, diese Einbußen auszugleichen?

- o Nein
- o Ja, nämlich
  - o Durch Ersparnisse
  - o Durch Kredit
  - o Unterstützung durch Familie/Freunde
  - o Verkauf/Beleihung von Eigentum
  - o Durch Sonstiges: _____
- o Dazu möchte ich keine Angabe machen

Müssen Sie sich aufgrund der Auswirkungen der Corona-Pandemie <u>zusätzlich</u> einschränken?

- o Nein
- o Ja
- o Dazu möchte ich keine Angabe machen

<u>V Informationen zu Ihrer Person</u>

Zum Schluss bitten wir Sie noch um einige Angaben zu Ihrer Person und zu Ihrer Lebenssituation, um Ihre Antworten besser einordnen zu können. Wir möchten Ihnen noch einmal versichern, dass alle Angaben völlig anonym sind und nicht Ihrer Person zugeordnet werden können.

Wie ist Ihre häusliche Situation?

- o Ich lebe zusammen mit meinem Ehepartner/meiner Ehepartnerin zusammen
- o Ich lebe mit meinem Lebenspartner/meiner Lebenspartnerin zusammen
- o Ich lebe ohne Partner/Partnerin
- o Anders: _____
- O Dazu möchte ich keine Angabe machen

Im Haushalt leben minderjährige Kinder: _____ (Anzahl)
Wie viele Personen leben derzeit insgesamt in Ihrem Haushalt? _____

10.5.4.2 Updated version of questionnaire after learnings from the pre-test.

I Krankheitssituation

Zunächst bitten wir Sie um einige Angaben zu Ihrer Erkrankungssituation.

Haben Sie eine bestätigte Krebsdiagnose?
O Nein -> *Ende der Befragung*
O Ja -> weiter

Welche Krebserkrankung wurde bei Ihnen zuletzt festgestellt?

- o Bauchspeicheldrüsenkrebs
- o Brustkrebs
- o Darmkrebs
- o Eierstockkrebs
- o Gallenblasenkrebs/Gallenwegskrebs
- o Gebärmutterhalskrebs
- o Gebärmutterkörperkrebs
- o Hautkrebs (nicht Melanom)
- o Hirntumor
- o Harnblasenkrebs
- o Hodenkrebs
- o Kehlkopfkrebs
- o Leberkrebs (NICHT Lebermetastasen)
- o Leukämie
- o Lungenkrebs
- o Magenkrebs
- o Malignes Melanom der Haut
- o Mesotheliom
- o Morbus Hodgkin
- o Mundhöhle/Rachen
- o Nierenkrebs
- o Non-Hodgkin-Lymphom
- o Plasmozytom (Multiples Myelom)
- o Prostatakrebs
- o Schilddrüsenkrebs
- o Speiseröhrenkrebs
- o Vulvakrebs
- o Weiß nicht
- o Sonstige (bitte notieren)

_____

Hat der Krebs gestreut? Liegen Metastasen in anderen Organen vor?

- o Nein
- o Ja
- o Verdacht
- o Weiß nicht

Wie ist Ihre Krankheitssituation? (Stadium/Phase)

- o Nach Diagnosestellung (Therapie hat noch nicht begonnen)
- o Während erster Behandlung, einschließlich ergänzender (adjuvanter) Therapie und Erhaltungstherapie
- o Erstbehandlung abgeschlossen
- o Rückfall/Rückfallbehandlung
- o Fortgeschrittene Erkrankung/palliative Behandlung
- o Weiß nicht

Welche Art von Behandlung oder Untersuchung wird gerade durchgeführt oder ist als nächstes geplant? *(Mehrfachantworten möglich)*

- o Operation zur Entfernung des Primärtumors
- o Operation zur Entfernung von Metastasen
- o Medikamentöse Therapie vor geplanter Operation (z.B. Chemotherapie, zielgerichtete Therapie, Immuntherapie)
- o Ergänzende (adjuvante) medikamentöse Therapie nach Operation (z.B. Chemotherapie, Hormontherapie, zielgerichtete Therapie, Immuntherapie)
- o Medikamentöse Therapie bei fortgeschrittener Erkrankung/bei Metastasen (z.B. Chemotherapie, Hormontherapie, zielgerichtete Therapie, Immuntherapie)
- o Bestrahlung der Tumorregion
- o Bestrahlung von Metastasen
- o Untersuchungen in der Nachsorge
- o Untersuchungen zur Verlaufskontrolle während einer Therapie (z. B. MRT, CT, Blutwerte)
- o Abwartendes Beobachten („Wait and See“)
- o Weiß nicht

In welchem Rahmen findet Ihre Behandlung bzw. Ihre Nachsorge bisher hauptsächlich statt?

- o Stationär im Krankenhaus
- o Ambulant im Krankenhaus
- o Onkologische Praxis
- o Andere (Fach)Arztpraxis *(bitte Fachrichtung angeben*):

_____

II Corona-Pandemie und Gesundheitsversorgung

Jetzt interessiert uns, ob die Corona-Pandemie Auswirkungen auf Ihre medizinische Versorgung hatte oder hat.

Sind Sie bzw. waren Sie selbst mit Sars-CoV2 (Coronavirus) infiziert oder an COVID-19 (Coronavirus-Erkrankung) erkrankt ?

- o Nein, ich bin weder aktuell noch war ich an der Coronavirus-Erkrankung erkrankt.
- o Ja, ich bin aktuell an Coronavirus-Erkrankung erkrankt.
- o Ja, ich war an Coronavirus-Erkrankung erkrankt und bin inzwischen wieder genesen.
- o Ich weiß nicht, ob ich mit Coronavirus infiziert bin oder war.

Ist oder war jemand in Ihrer Familie oder Ihrem Freundeskreis an Coronavirus-Erkrankung erkrankt oder mit Coronavirus infiziert? *(Mehrfachnennungen möglich)*

- o Ja, in meiner Familie
- o Ja, in meinem engen Freundeskreis
- o Ja, sonstige Personen in deren Nähe ich mich aufhielt (z.B. Pflegeheim/betreutes Wohnen, Reha, Nachbarn)
- o Nein bzw. weiß nicht

Hat sich durch die Corona-Pandemie etwas in Ihrem Behandlungs- oder Nachsorgeplan geändert?

- o Nein, es hat sich nichts geändert *Falls nein: Weiter bei Sektion 3*
- o Ja, es hat sich etwas geändert

Falls Ja wird die folgende Liste von Prozeduren/Maßnahmen angeboten, aus der eine oder mehrere angekreuzt werden können. Die Folgefragen erscheinen dann pro ausgewählter Änderung.

Bitte wählen Sie aus, bei welcher Maßnahme/bei welchen Maßnahmen eine Änderung erfolgt ist (Mehrfachnennungen möglich).

- o Operation
- o Medikamentöse Tumortherapie (z. B. Chemotherapie, Hormontherapie, zielgerichtete Therapie, Immuntherapie)
- o Bestrahlung
- o Verlaufskontrolle während der Behandlung
- o Nachsorgetermin
- o Psychosoziale / psychoonkologische Beratung oder Betreuung
- o Betreuung durch Pflegedienst

- o **Operation**

  - o Verschoben *Bei Ankreuzen erscheinen die folgenden 6 Fragen*
  - o Andere Behandlungsmethode *Danach erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (nicht Verschiebung)*

Was war der Grund für die Änderung? *(Mehrfachnennungen möglich)*
o Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
o Die Gefahr, dass ich Andere mit dem Corona-Virus anstecken könnte
o Versorgungsengpass wegen der Corona-Pandemie
o Meine gesundheitliche Verfassung
o Die Operation ist nicht dringend
o Weiß nicht/wurde mir nicht mitgeteilt
o Anderer Grund: _____

Für wie lange wurde verschoben?
o Bis 4 Wochen
o Länger als 4 Wochen
o Auf unbestimmte Zeit

Wer hat die Entscheidung getroffen?
o Ich selbst
o Die behandelnden Ärzte
o Die behandelnden Ärzte und ich gemeinsam

Wie sehr belastet Sie diese Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

Haben Sie wegen dieser Änderung professionelle Unterstützung / Beratung gesucht?
o Nein *->Weiter zur nächsten Frage – „Was denken Sie selbst über die Änderungen..."*
o Ja *-> Kategorien erscheinen*

Bitte Zutreffendes auswählen; Mehrfachnennungen möglich
o Durch meinen behandelnden Arzt
o Durch einen Psycho-Onkologen
o Durch einen anderen Arzt / eine andere Klinik
o Durch meine Krankenversicherung
o Durch eine Beratungsstelle / Beratungsdienst
o Anderes *(bitte angeben)*: _____

Was denken Sie selbst über diese Änderung bei Ihrer Behandlung oder Versorgung?

	Stimme voll und ganz zu	Stimme zu	Weder noch	Stimme nicht zu	Stimme gar nicht zu	Kann ich nicht beurteilen/ weiß nicht
Sie erscheint mir angemessen	o	o	o	o	o	o
Sie beeinträchtigt (möglicherweise) den Erfolg meiner Behandlung	o	o	o	o	o	o
Ich habe durch die Änderung weitere Nachteile	o	o	o	o	o	o

o **Medikamentöse Tumortherapie (z. B. Chemotherapie, Hormontherapie, zielgerichtete Therapie, Immuntherapie)**

- o Verschoben *Bei Ankreuzen erscheinen die folgenden 6 Fragen*
- o Entfällt *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (nicht Verschiebung)*
- o Andere Therapie *Danach erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (nicht Verschiebung)*

Was war der Grund für die Änderung? *(Mehrfachnennungen möglich)*

- o Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
- o Versorgungsengpass wegen der Corona-Pandemie
- o Meine gesundheitliche Verfassung
- o Die Therapie ist nicht dringend
- o Weiß nicht/wurde mir nicht mitgeteilt
- o Anderer Grund: _____

Für wie lange wurde verschoben?

- o Bis 4 Wochen
- o Länger als 4 Wochen
- o Auf unbestimmte Zeit

Wer hat die Entscheidung getroffen?

- o Ich selbst
- o Die behandelnden Ärzte
- o Die behandelnden Ärzte und ich gemeinsam

Wie sehr belastet Sie diese Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

Haben Sie wegen dieser Änderung professionelle Unterstützung / Beratung gesucht?

o Nein *->Weiter zur nächsten Frage – „Was denken Sie selbst über die Änderungen..."*

o Ja *-> Kategorien erscheinen*

Bitte Zutreffendes auswählen; Mehrfachnennungen möglich

o Durch meinen behandelnden Arzt
o Durch einen Psycho-Onkologen
o Durch einen anderen Arzt / eine andere Klinik
o Durch meine Krankenversicherung
o Durch eine Beratungsstelle / Beratungsdienst
o Anderes (*bitte angeben*): _____

Was denken Sie selbst über diese Änderung bei Ihrer Behandlung oder Versorgung?

	Stimme voll und ganz zu	**Stimme zu**	**Weder noch**	**Stimme nicht zu**	**Stimme gar nicht zu**	**Kann ich nicht beurteilen/ weiß nicht**
Sie erscheint mir angemessen	O	O	O	O	O	O
Sie beeinträchtigt (möglicherweise) den Erfolg meiner Behandlung	O	O	O	O	O	O
Ich habe durch die Änderung weitere Nachteile	O	O	O	O	O	O

o **Bestrahlung**

o Verschoben *Bei Ankreuzen erscheinen die folgenden 6 Fragen*
o Entfällt *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (nicht Verschiebung)*
o Andere Therapie *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (nicht Verschiebung)*

Was war der Grund für die Änderung? *(Mehrfachnennungen möglich)*
- o Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
- o Versorgungsengpass wegen der Corona-Pandemie
- o Meine gesundheitliche Verfassung
- o Die Bestrahlung ist nicht dringend
- o Weiß nicht/wurde mir nicht mitgeteilt
- o Anderer Grund: _____

Für wie lange wurde verschoben?
- o Bis 4 Wochen
- o Länger als 4 Wochen
- o Auf unbestimmte Zeit

Wer hat die Entscheidung getroffen?
- o Ich selbst
- o Die behandelnden Ärzte
- o Die behandelnden Ärzte und ich gemeinsam

Wie sehr belastet Sie diese Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

Haben Sie wegen dieser Änderung professionelle Unterstützung / Beratung gesucht?

o Nein *->Weiter zur nächsten Frage – „Was denken Sie selbst über the Änderungen...“*

o Ja *-> Kategorien erscheinen*

Bitte Zutreffendes auswählen; Mehrfachnennungen möglich
- o Durch meinen behandelnden Arzt
- o Durch einen Psycho-Onkologen
- o Durch einen anderen Arzt / eine andere Klinik
- o Durch meine Krankenversicherung
- o Durch eine Beratungsstelle / Beratungsdienst
- o Anderes *(bitte angeben)*:

_____

Was denken Sie selbst über diese Änderung bei Ihrer Behandlung oder Versorgung?

	Stimme voll und ganz zu	Stimme zu	Weder noch	Stimme nicht zu	Stimme gar nicht zu	Kann ich nicht beurteilen/ weiß nicht
Sie erscheint mir angemessen	O	O	O	O	O	O
Sie beeinträchtigt (möglicherweise) den Erfolg meiner Behandlung	O	O	O	O	O	O
Ich habe durch die Änderung weitere Nachteile	O	O	O	O	O	O

o **Verlaufskontrolle während der Behandlung**

- o Verschoben *Bei Auswahl erscheinen die folgenden 6 Fragen*
- o Entfällt *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (nicht Verschiebung)*
- o Andere Form des Kontakts (Telefon, Videochat, Skype…) *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (nicht Verschiebung)*

Was war der Grund für die Änderung? *(Mehrfachnennungen möglich)*
- o Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
- o Versorgungsengpass wegen der Corona-Pandemie
- o Meine gesundheitliche Verfassung
- o Die Untersuchungen sind nicht dringend
- o Weiß nicht / wurde mir nicht mitgeteilt
- o Anderer Grund: _____

Für wie lange wurde verschoben?
- o Bis 4 Wochen
- o Länger als 4 Wochen
- o Auf unbestimmte Zeit

Wer hat die Entscheidung getroffen?
- o Ich selbst
- o Die behandelnden Ärzte
- o Die behandelnden Ärzte und ich gemeinsam

Wie sehr belastet Sie diese Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

Haben Sie wegen dieser Änderung professionelle Unterstützung / Beratung gesucht?

O Nein *->Weiter zur nächsten Frage – „Was denken Sie selbst über the Änderungen...“*

O Ja *-> Kategorien erscheinen*

Bitte Zutreffendes auswählen; Mehrfachnennungen möglich

O Durch meinen behandelnden Arzt
O Durch einen Psycho-Onkologen
O Durch einen anderen Arzt / eine andere Klinik
O Durch meine Krankenversicherung
O Durch eine Beratungsstelle / Beratungsdienst
O Anderes (*bitte angeben*): _____

Was denken Sie selbst über diese Änderung bei Ihrer Behandlung oder Versorgung?

	Stimme voll und ganz zu	**Stimme zu**	**Weder noch**	**Stimme nicht zu**	**Stimme gar nicht zu**	**Kann ich nicht beurteilen/ weiß nicht**
Sie erscheint mir angemessen	O	O	O	O	O	O
Sie beeinträchtigt (möglicherweise) den Erfolg meiner Behandlung	O	O	O	O	O	O
Ich habe durch die Änderung weitere Nachteile	O	O	O	O	O	O

o **Nachsorgetermin**

o Verschoben *Bei Ankreuzen erscheinen die folgenden 6 Fragen*
o Entfällt *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (nicht Verschiebung)*
o Andere Form des Kontakts (Telefon, Videochat, Skype…) *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (nicht Verschiebung)*

Was war der Grund für die Änderung? *(Mehrfachnennungen möglich)*
o Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
o Versorgungsengpass wegen der Corona-Pandemie
o Meine gesundheitliche Verfassung
o Der Nachsorgetermin ist nicht dringend
o Weiß nicht / wurde mir nicht mitgeteilt
o Anderer Grund: _____

Für wie lange wurde verschoben?
o Bis 4 Wochen
o Länger als 4 Wochen
o Auf unbestimmte Zeit

Wer hat die Entscheidung getroffen?
o Ich selbst
o Die behandelnden Ärzte
o Die behandelnden Ärzte und ich gemeinsam

Wie sehr belastet Sie die Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

Haben Sie wegen dieser Änderung professionelle Unterstützung / Beratung gesucht?
o Nein *->Weiter zur nächsten Frage – „Was denken Sie selbst über the Änderungen..."*
o Ja *-> Kategorien erscheinen*

Bitte Zutreffendes auswählen; Mehrfachnennungen möglich
o Durch meinen behandelnden Arzt
o Durch einen Psycho-Onkologen
o Durch einen anderen Arzt / eine andere Klinik
o Durch meine Krankenversicherung
o Durch eine Beratungsstelle / Beratungsdienst
o Anderes (bitte angeben): _____

Was denken Sie selbst über diese Änderung bei Ihrer Behandlung oder Versorgung?

	Stimme voll und ganz zu	Stimme zu	Weder noch	Stimme nicht zu	Stimme gar nicht zu	Kann ich nicht beurteilen/ weiß nicht
Sie erscheint mir angemessen	O	O	O	O	O	O
Sie beeinträchtigt (möglicherweise) den Erfolg meiner Behandlung	O	O	O	O	O	O
Ich habe durch die Änderung weitere Nachteile	O	O	O	O	O	O

O **Psychosoziale oder psychoonkologische Beratung / Betreuung**

- O Verschoben *Bei Auswahl erscheinen die folgenden 6 Fragen*
- O Entfällt *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (nicht Verschiebung)*
- O Andere Form des Kontakts (Telefon, Videochat, Skype...) *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (nicht Verschiebung)*

Was war der Grund für die Änderung? *(Mehrfachnennungen möglich)*
- O Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
- O Versorgungsengpass wegen der Corona-Pandemie
- O Meine gesundheitliche Verfassung
- O Die Beratung ist nicht dringend
- O Weiß nicht / wurde mir nicht mitgeteilt
- O Anderer Grund: _____

Für wie lange wurde verschoben?
- O Bis 4 Wochen
- O Länger als 4 Wochen
- O Auf unbestimmte Zeit

Wer hat die Entscheidung getroffen?
- O Ich selbst
- O Meine Betreuer/Therapeuten
- O Meine Betreuer/Therapeuten und ich gemeinsam

Wie sehr belastet Sie die Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

Haben Sie wegen dieser Änderung professionelle Unterstützung / Beratung gesucht?

O Nein *->Weiter zur nächsten Frage – „Was denken Sie selbst über the Änderungen..."*

O Ja *-> Kategorien erscheinen*

Bitte Zutreffendes auswählen; Mehrfachnennungen möglich

O Durch meinen behandelnden Arzt
O Durch einen Psycho-Onkologen
O Durch einen anderen Arzt / eine andere Klinik
O Durch meine Krankenversicherung
O Durch eine Beratungsstelle / Beratungsdienst
O Anderes (bitte angeben):

_____

Was denken Sie selbst über diese Änderung bei Ihrer Behandlung oder Versorgung?

	Stimme voll und ganz zu	**Stimme zu**	**Weder noch**	**Stimme nicht zu**	**Stimme gar nicht zu**	**Kann ich nicht beurteilen/ weiß nicht**
Sie erscheint mir angemessen	O	O	O	O	O	O
Sie beeinträchtigt (möglicherweise) den Erfolg meiner Behandlung	O	O	O	O	O	O
Ich habe durch die Änderung weitere Nachteile	O	O	O	O	O	O

O **Betreuung durch Pflegedienst**

- O Verschoben *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (**nicht Reduzierung**)*
- O Entfällt *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (**nicht Verschiebung, nicht Reduzierung**)*
- O Reduziert *Bei Ankreuzen erscheinen die Fragen zu Grund, Reduzierung, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (**nicht Verschiebung**)*
- O Umgestellt auf andere Form des Kontakts (Telefon, Videochat, Skype …) *Bei Ankreuzen erscheinen die Fragen zu Grund, Entscheidung, Belastung, Suche nach Unterstützung / Beratung, und Gedanken über Änderungen (**nicht Verschiebung, nicht Reduzierung**)*

Was war der Grund für die Änderung? *(Mehrfachnennungen möglich)*
- O Die Gefahr, dass ich mich mit dem Corona-Virus anstecken könnte
- O Versorgungsengpass wegen der Corona-Pandemie
- O Meine gesundheitliche Verfassung
- O Die Betreuung ist nicht dringend
- O Weiß nicht / wurde mir nicht mitgeteilt
- O Anderer Grund: _____

Für wie lange wurde verschoben?
- O Bis 4 Wochen
- O Länger als 4 Wochen
- O Auf unbestimmte Zeit

Um wie viele Stunden pro Woche ist die pflegerische Betreuung reduziert?
- O 0 – 1 Stunde weniger pro Woche
- O 1 – 5 Stunden weniger pro Woche
- O 5 – 10 Stunden weniger pro Woche
- O Mehr als 10 Stunden weniger pro Woche

Wer hat die Entscheidung getroffen?
- O Ich selbst
- O Meine Betreuer/mein Arzt
- O Meine Betreuer/ mein Arzt und ich gemeinsam

Wie sehr belastet Sie die Änderung? *Bitte wählen Sie aus von 0 (belastet mich gar nicht) bis 10 (belastet mich extrem).*

Belastet mich gar nicht ----------> Belastet mich extrem										
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

Haben Sie wegen dieser Änderung professionelle Unterstützung / Beratung gesucht?

o Nein *->Weiter zur nächsten Frage – „Was denken Sie selbst über the Änderungen...“*

o Ja *-> Kategorien erscheinen*

Bitte Zutreffendes auswählen; Mehrfachnennungen möglich

o Durch meinen behandelnden Arzt
o Durch einen Psycho-Onkologen
o Durch einen anderen Arzt / eine andere Klinik
o Durch meine Krankenversicherung
o Durch eine Beratungsstelle / Beratungsdienst
o Anderes (*bitte angeben*): _____

Was denken Sie selbst über diese Änderung bei Ihrer Behandlung oder Versorgung?

	Stimme voll und ganz zu	**Stimme zu**	**Weder noch**	**Stimme nicht zu**	**Stimme gar nicht zu**	**Kann ich nicht beurteilen/ weiß nicht**
Sie erscheint mir angemessen	O	O	O	O	O	O
Sie beeinträchtigt (möglicherweise) den Erfolg meiner Behandlung	O	O	O	O	O	O
Ich habe durch die Änderung weitere Nachteile	O	O	O	O	O	O

III Lebensqualität und Belastung

Jetzt würden wir gerne erfahren, welche Gedanken und Sorgen Sie als Folge der Corona-Pandemie in Bezug auf Ihre persönliche Situation bewegen.

Bitte bewerten Sie für den Zeitraum der letzten 4 Wochen, inwieweit die folgenden Aussagen für Sie zutreffen.

Ich mache mir Sorgen...	**trifft ganz genau zu**	**trifft eher zu**	**teils-teils**	**trifft eher nicht zu**	**trifft gar nicht zu**
... um meine Arbeit und mögliche finanzielle Folgen.	O	O	O	O	O
... wegen finanzieller Mehrbelastungen durch meine Krebserkrankung.	O	O	O	O	O
... dass ich für längere Zeit finanzielle Einbußen haben werde.	O	O	O	O	O
... dass ich selbst an der Coronavirus-Erkrankung erkranken oder sterben könnte.	O	O	O	O	O
… dass Angehörige/Freunde an der Coronavirus-Erkrankung erkranken oder sterben könnten	O	O	O	O	O
... wegen möglicher Auswirkungen auf die Qualität meiner medizinischen Versorgung.	O	O	O	O	O

Wie erleben Sie die Einschränkungen persönlicher Begegnungen durch die Corona-Pandemie? *(Bitte bewerten Sie, wie sehr Ihnen die verschiedenen Kontakte fehlen.)*

	fehlen mir sehr	**fehlen mir eher**	**teils-teils**	**fehlen mir eher nicht**	**fehlen mir gar nicht**
Persönliche Kontakte mit Angehörigen, Freunden, Arbeitskollegen, Nachbarn …	O	O	O	O	O
Persönliche Kontakte mit anderen Krebspatienten/mit meiner Selbsthilfegruppe …	O	O	O	O	O
Persönliche Kontakte mit Ärzten und Pflegenden …	O	O	O	O	O
Persönliche Kontakte mit Betreuern, Therapeuten und anderen Helfern …	O	O	O	O	O
Persönliche Kontakte in der Öffentlichkeit (Lokalbesuch, Park, Konzert, Theater, Einkaufen …) …	O	O	O	O	O

Was hilft Ihnen mit der aktuellen Situation umzugehen? *(Mehrfachnennungen möglich)*

O Ich fühle mich durch die aktuelle Situation nicht beeinträchtigt. *Weiter zu Sektion IV*

o Ich nutze professionelle Unterstützung (Psychologe, Beratungsstelle …).
o Ich telefoniere/chatte/benutze Soziale Medien mit Familie, Freunden und Bekannten.
o Ich finde Frieden in der Natur/Gartenarbeit.
o Mein Haustier/meine Haustiere bringen mir Trost und Freude.
o Ich finde Trost in meiner Religion.
o Mein Optimismus und meine Zuversicht geben mir Kraft.
o Ich betreibe Sport, Entspannungsübungen oder Meditation.
o Ich schaue Filme und Serien, lese, höre Musik.
o Ich betreibe Sport, Entspannungsübungen oder Meditation.
o Anderes *(bitte angeben)*: _____

IV Finanzielle Folgen der Krebserkrankung und der Corona-Pandemie

Neben den körperlichen und seelischen Belastungen kann eine Krebserkrankung auch Auswirkungen auf die finanzielle Situation haben. Die Corona-Pandemie könnte zusätzliche Kosten oder Einbußen verursachen. Wir möchten erfahren, wie das bei Ihnen ist.

Wie ist Ihr derzeitiger Erwerbsstatus?

o Angestellt
o Selbständig oder freiberuflich tätig
o Beamter/Beamtin
o Berentet
O Arbeitslos *weiter zu Frage 16*
o Nicht erwerbstätig *weiter zu Frage 16*
o Möchte ich nicht beantworten *weiter zu Frage 16*

In welchem Umfang sind Sie angestellt?
o Vollzeit
o Teilzeit
o Minijob

In den nächsten Fragen geht es um mögliche Folgen Ihrer Krebserkrankung.

Waren Sie aufgrund Ihrer Krebserkrankung in den letzten 4 Wochen arbeitsunfähig?

o Nein *-> weiter zu Frage 17*
o Ja

Wie lange? *(bitte der Gesamtzahl der bisherigen Arbeitsunfähigkeits-Tage in den letzten 4 Wochen angeben*) _____

Haben oder hatten Sie <u>infolge Ihrer Krebserkrankung</u> zusätzliche Ausgaben im Zusammenhang mit Ihrer Gesundheitsversorgung?

o Nein *-> weiter bei Frage 18*
o Ja *-> Folgefrage erscheint „Welche waren/sind das?“ „Wie hoch...“*
o Dazu möchte ich keine Angabe machen *-> weiter bei Frage 18*

Welche waren/sind das? *(Mehrfachnennungen möglich)*

- o Zuzahlungen für ambulante oder stationäre Leistungen
- o Fahrtkosten
- o Inanspruchnahme von Haushaltshilfen
- o Behandlungen und Medikamente, die nicht von der Krankenkasse erstattet werden
- o Pflegekraft/Pflegedienst
- o Sonstiges:_____

Wie hoch waren/sind diese Mehrausgaben etwa monatlich?

- o Unter 100 Euro
- o 100 – 200 Euro
- o 201 – 500 Euro
- o 501 – 800 Euro
- o 801 – 1.200 Euro
- o Über 1.200 Euro
- o Dazu möchte ich keine Angabe machen
- O Weiß nicht

Hat <u>die Krebserkrankung</u> dazu geführt, dass Sie im Alltag mehr sparen müssen?

O Nein *-> weiter bei Frage 19*
o Ja *-> weiter bei Folgefrage „Wo versuchen Sie Geld...“*
o Dazu möchte ich keine Angaben machen *-> weiter bei Frage 19*

Wo versuchen Sie Geld einzusparen? *(Mehrfachnennungen möglich)*

- O Bei medizinischen Behandlungen/Zusatzleistungen
- O In meiner Freizeitgestaltung (z. B. Kinobesuche, Essen-/Ausgehen, Reisen, Sport)
- O Bei Frisör, Pflege, Kosmetik
- O Bei meiner Ernährung/bei Lebensmitteln
- O Beim Kauf von Bekleidung
- O Bei Genussmitteln (Tabak, Alkohol …)
- O Bei Ausstattung für Haushalt und Wohnen
- O Sonstiges: _____

Mögliche <u>zusätzliche</u> Auswirkungen der Corona-Pandemie

Sind Ihnen <u>in den letzten 4 Wochen</u> durch die Coronavirus-Pandemie zusätzliche Kosten für Ihre Gesundheitsversorgung entstanden?

- o Nein *-> weiter bei Frage 20*
- o Ja *-> weiter bei Folgefrage „Haben Sie Leistungen..."*
- o Dazu möchte ich keine Angabe machen *-> weiter bei Frage 20*

Haben Sie Leistungen <u>auf eigene Kosten</u> in Anspruch genommen oder gesundheitliche Ausgaben selbst getragen?

- O Nein
- o Ja *-> Folgefragen erscheinen: „Welche waren das?" „Wie viel Geld..."*
- O Dazu möchte ich keine Angabe machen

Welche waren das? *(Mehrfachnennungen möglich)*

- o Private medizinische Behandlung
- o Komplementäre oder alternative Behandlungsmethoden
- o Labordiagnostik/apparative Untersuchungen
- O Kauf von Masken, Desinfektionsmittel etc.
- o Andere *(Bitte angeben)*: _____

Wie viel Geld haben Sie für diese Leistungen <u>in den letzten 4 Wochen</u> etwa ausgegeben?

- o Unter 100 Euro
- o 100 – 200 Euro
- o 201 – 500 Euro
- o 501 – 800 Euro
- o 801 – 1.200 Euro
- o Über 1.200 Euro
- o Dazu möchte ich keine Angabe machen

Sind Ihnen durch die Corona-Pandemie <u>Einkommenseinbußen</u> entstanden, z.B. durch den Verlust eines (Neben-)Jobs, Verringerung der Arbeitszeit, Kurzarbeit, Verlust von Mieteinnahmen o.ä.?

- o Nein *-> weiter bei Frage 21*
- o Ja *-> Folgefrage erscheint*

Die Einbußen betrugen monatlich netto

- o Unter 100 Euro
- o 100-200 Euro
- o 201-500 Euro
- o 501-800 Euro
- o 801-1.200 Euro
- o 1.200-1.600 Euro
- o 1.600 – 2.000 Euro
- o Über 2.000 Euro
- o Dazu möchte ich keine Angabe machen *-> weiter bei Frage 21*

Haben/hatten Sie die Möglichkeit, diese Einbußen aufzufangen oder zu überbrücken?

- o Nein -> *weiter bei Frage 21*
- o Ja -> *Folgefrage erscheint*

Wie? *(Mehrfachnennungen möglich)*

- o Durch Ersparnisse
- o Durch Kredit
- o Unterstützung durch Familie/Freunde
- o Verkauf/Beleihung von Eigentum
- o Durch Einsparung von Fahrtkosten wegen Arbeit im Homeoffice
- o Dazu möchte ich keine Angabe machen
- o Anders: _____

Müssen Sie sich aufgrund der Auswirkungen der Corona-Pandemie <u>zusätzlich</u> finanziell einschränken?

- o Nein -> *weiter bei Frage 22*
- o Ja -> *Folgefrage erscheint*
- o Dazu möchte ich keine Angabe machen -> *weiter bei Frage 22*

Woran sparen Sie (*Mehrfachangaben möglich)*

- o Bei medizinischen Behandlungen/Zusatzleistungen
- o In meiner Freizeitgestaltung (z. B. Kinobesuche, Essen-/Ausgehen, Reisen, Sport)
- o Bei Frisör, Pflege, Kosmetik
- o Bei meiner Ernährung/bei Lebensmitteln
- o Beim Kauf von Bekleidung
- o Bei Genussmitteln (Tabak, Alkohol …)
- o Bei Ausstattung für Haushalt und Wohnen
- o Sonstiges: _____

V Informationen zu Ihrer Person

Zum Schluss bitten wir Sie noch um einige Angaben zu Ihrer Person und zu Ihrer Lebenssituation, um Ihre Antworten besser einordnen zu können. Wir möchten Ihnen noch einmal versichern, dass alle Angaben völlig anonym sind und nicht Ihrer Person zugeordnet werden können.

In welchem Jahr sind Sie geboren?

_____ *(bitte Geburtsjahr angeben)*

Geschlecht

o Männlich
o Weiblich
O Sonstige

Ihr höchster Schulabschluss

o Hauptschulabschluss/Volksschule
o Fachoberschulreife/Realschulabschluss
o Fachhochschulreife/Fachabitur
o Hochschulreife/Abitur
o Kein Abschluss
o Dazu möchte ich keine Angabe machen

Wie ist Ihre häusliche Situation?

o Ich lebe zusammen mit meinem Ehepartner/meiner Ehepartnerin zusammen
o Ich lebe mit meinem Lebenspartner/meiner Lebenspartnerin zusammen
o Ich lebe ohne Partner/Partnerin
O Ich lebe mit Partner/Partnerin, aber jeweils mit eigenem Haushalt
O Ich lebe in einer WG
O Ich lebe mit Familienangehörigen
o Dazu möchte ich keine Angabe machen
o Anders: _____

Wie viele Personen (Sie selbst eingeschlossen) leben derzeit insgesamt in Ihrem Haushalt? *(Bitte Zahl angeben, „1" falls Sie alleine leben)* _____

Wie viele minderjährige Kinder leben im Haushalt? *(Bitte Zahl angeben, ggf. „0")*: _____

Wie sind Sie krankenversichert?

o Privat
o Gesetzlich
o Gesetzlich plus private Zusatzversicherung
o Beitragsfrei mitversichert
o Dazu möchte ich keine Angabe machen
o Anders, und zwar:_____

Wie hoch ist das gesamte verfügbare monatliche Netto-Einkommen Ihres Haushalts?

- o Unter 1.200 Euro
- O 1.201 – 2.000 Euro
- o 2.001 – 3.000 Euro
- o 3.001 – 4.000 Euro
- o 4.001 – 5.000 Euro
- o Über 5.000 Euro
- o Dazu möchte ich keine Angabe machen

11 CURRICULUM VITAE

PERSONAL DATA

Name	Ubels, Jasper Tobias
Date of Birth:	06-07-1993
Place of Birth:	Den Helder

SCHOOL EDUCATION

2005 – 2011	Regius College Schagen, Schagen, Netherlands
16.06.2011	Gymnasium Diploma

UNIVERSITY EDUCATION

2011-2015	Vrije Universiteit Amsterdam, Amsterdam, Netherlands
15.08.2015	Bachelor of Science in Health and Life sciences
2015-2017	Umeå Universitet, Umeå, Sweden
23.08.2017	Master of Science in Public Health, Specialization in Health Economics Master Theses: 1. Setting Cost-Effectiveness Thresholds: Clashing Opinions, Ignorance, and an Absence of Data. 2. Screening for Rheumatic Heart Disease: A Markov Model
2019-Now	Doctoral Student Division of Health Economics, Deutsches Krebsforschungszentrum (DKFZ), Germany

PROFESSIONAL EXPERIENCE

2017-2019	Researcher, Division of Health Economics, Deutsches Krebsforschungszentrum (DKFZ), Germany

12 PUBLICATIONS

PUBLICATIONS RELATED TO DISSERTATION

PEER-REVIEWED PUBLICATIONS

Ubels, J., Hernandez-Villafuerte, K., and Schlander, M. (2022). The value of freedom: A review of the current developments and conceptual issues in the measurement of capability. Journal of human development and capabilities 23, 327-353.

Linton, M.J., Mitchell, P.M., Al-Janabi, H., Schlander, M., Richardson, J., Iezzi, A., Ubels, J., and Coast, J. (2020). Comparing the German translation of the ICECAP-A capability wellbeing measure to the original English version: Psychometric properties across healthy samples and seven health condition groups. Applied Research in Quality of Life 15, 651-673.

Gaisser, A., Eckford, R.D., Arndt, V., Doege, D., Kludt, E., Ubels, J., Schlander, M., and Weg-Remers, S. (2022). Nearly two years of the coronavirus pandemic from the perspective of people affected by cancer. Der Onkologe: Organ der Deutschen Krebsgesellschaft eV, 1-5.

Eckford, R.D., Gaisser, A., Arndt, V., Baumann, M., Kludt, E., Mehlis, K., Ubels, J., Winkler, E.C., Weg-Remers, S., and Schlander, M. (2022). The COVID-19 pandemic and cancer patients in Germany: impact on treatment, follow-up care and psychological burden. Frontiers in Public Health 9, 2364.

NON-PEER-REVIEWED PUBLICATION

Book chapter:
Ubels, J. (2021). The assessment of value in health economics: utility and capability. In Defining the Value of Medical Interventions: Normative and Empirical Challenges C. Buch, J. Schildmann, and J. Zerth, eds. (Stuttgart: Kohlhammer Verlag), pp. 69 - 82.

OTHER PUBLICATIONS

PEER-REVIEWED PUBLICATIONS

Pham, P. D., Schlander, M., Eckford, R., Hernandez-Villafuerte, K., Ubels, J. (2023). Developing a conceptual framework for socioeconomic impact research in European cancer patients. The patient (in press).

Ubels, J., Sable, C., Beaton, A.Z., Nunes, M.C.P., Oliveira, K.K., Rabelo, L.C., Teixeira, I.M., Ruiz, G.Z., Rabelo, L.M.M., Tompsett, A.R. and Ribeiro, A.L.P., 2020. Cost-effectiveness of rheumatic heart disease echocardiographic screening in Brazil: Data from the PROVAR+ study. Global heart, 15(1).

Ran, T., Cheng, C.-Y., Misselwitz, B., Brenner, H., Ubels, J., and Schlander, M. (2019). Cost-effectiveness of colorectal cancer screening strategies—a systematic review. Clinical gastroenterology and hepatology 17, 1969-1981. e1915.

Cameron, D.*, Ubels, J.*, and Norström, F. (2018). On what basis are medical cost-effectiveness thresholds set? Clashing opinions and an absence of data: a systematic review. Global health action 11, 1447828.

* Shared first-authorship

NON-PEER-REVIEWED PUBLICATION

Editorial:
Ubels, J. (2021). Saving money, saving lives: Health economics as a guide to cost-effective decision making. International journal of cardiology 339, 130-131.

13 ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor, Prof. Schlander, for giving me the opportunity to work on the project presented in this dissertation and for his invaluable guidance. Thanks to his belief, trust, and insights, I have grown not only professionally, but also personally.

I would also like to express my gratitude to Prof. Holm, Prof. Richardson, and Dr. Hernández-Villafuerte. Their contributions and insights enriched this dissertation immeasurably. Furthermore, I would like to thank Prof. Hagemann for providing indispensable advice during a crucial moment of my PhD project.

Additionally, I would like to thank all my colleagues in the department for providing an inspiring but more importantly a fun environment to work in. I learned a lot from them and enjoyed the time we spend together.

Finally, I would like to thank Julia, my girlfriend, for her patience and support. I am lucky to have her in my life.