



Semantic primes theory may be helpful in designing questionnaires such as to prevent response shift

Antoine Vanier, Alain Leplège, Jean-Benoit Hardouin, Véronique Sébille,
Bruno Falissard

► To cite this version:

Antoine Vanier, Alain Leplège, Jean-Benoit Hardouin, Véronique Sébille, Bruno Falissard. Semantic primes theory may be helpful in designing questionnaires such as to prevent response shift. *Journal of Clinical Epidemiology*, Elsevier, 2015, 68 (6), pp.646-654. <10.1016/j.jclinepi.2015.01.023>. <hal-01139064>

HAL Id: hal-01139064

<http://hal.upmc.fr/hal-01139064>

Submitted on 7 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TITLE: Semantic primes theory may be helpful in designing questionnaires such as to prevent response shift

Antoine Vanier^{1,2,3,4*}, Alain Leplège⁵, Jean-Benoit Hardouin², Véronique Sébille², Bruno Falissard^{1,6,7}

1 INSERM, U669 PSIGIAM, Paris, France

2 University of Nantes, EA4275 “Biostatistics, Pharmacoepidemiology and Subjective Measures in Health Sciences”, France

3 UPMC Univ. Paris 06, Department of Biostatistics, Paris, France

4 AP-HP, University Hospitals Pitié-Salpêtrière Charles-Foix, Department of Biostatistics Public Health and Medical Informatics, Paris, France

5 Univ. Paris Diderot, Sorbonne Paris Cité, Department of History and Philosophy of Sciences, Paris, France

6 Univ. Paris-Sud and Univ. Paris-Descartes, Paris, France

7 AP-HP, University Hospitals Paris-Sud, Department of Public Health, Le Kremlin-Bicêtre, France

***Corresponding author:** Antoine Vanier

Dpt de Biostatistique Santé Publique et Information Médicale

AP-HP Hôpital de la Pitié-Salpêtrière

47-83 Boulevard de l'hôpital

75651 Paris Cedex 13, France

Tel: (33) 1 42 16 02 67

Mail: antoine.vanier@psl.aphp.fr

DISCLOSURE OF FINANCIAL CONFLICT OF INTEREST

Antoine Vanier: None

Alain Leplège: None

Jean-Benoit Hardouin: None

Véronique Sébille: None

Bruno Falissard: None

ACCEPTED MANUSCRIPT

ABSTRACT

Objective. The purpose of Randomized Control Trials (RCTs) can be the assessment of the direct effect of treatment on Health-Related Quality of Life (HRQL). Response shift (RS) theory considers that a change in HRQL scores observed over time cannot be explained solely by a direct effect of a medical condition, it may also result from a change in the way people appraise their HRQL. The RS effect is a potential bias that is liable to compromise efficient assessment of the effect of treatment on HRQL.

Study design and setting. We hypothesize a link between the RS effect on HRQL scores, and the level of complexity of HRQL conceptualization.

Results. We discuss how the impact of reconceptualization on scores depends on the complexity of the linguistic definition of a subjective construct, and how for reprioritization the impact depends on the dimensionality. The linguistic theory of semantic primes is used to help identify how subjective constructs can be classified according to the complexity of their definitions.

Conclusion. Finally, we suggest that the impact of the RS effect on HRQL scores could be avoided (or lessened) if questionnaires were designed with a rule of “the least semantic and psychometric complexity” in mind.

Key words: Health-Related Quality of Life, Patient-Reported Outcomes, Response Shift, Methodology, Randomized Control Trials, Psychometrics

Running Head: Links between complexity (in semantic and psychometric terms) and response shift

Word count: 4000 (Main manuscript), 199 (Abstract)

What is new?**- What is known:**

- Patient-Reported Outcomes (PRO), especially those assessing Health-Related Quality of Life (HRQL), are increasingly used as endpoints in Randomized Control Trials (RCTs) to assess the direct effect of a treatment on HRQL.
- Response shift (RS) theory considers that a change in HRQL scores observed over time cannot be explained solely by the direct effect of a medical condition, it can also result from a change in the way people appraise their HRQL.
- Therefore, the RS effect can be a bias, liable to compromise efficient assessment of treatment effect on HRQL.

- What this adds to what is known:

- We hypothesize a link between the RS effect on HRQL scores, and the level of complexity of HRQL conceptualization.
- We postulate that the linguistic theory of semantic primes can be used to categorize subjective construct definitions in terms of low or high semantic complexity.
- We hypothesize that the impact of reconceptualization on HRQL scores differs according to semantic complexity, and that for reprioritization, it depends on dimensionality (in psychometric terms).

- What is the implication, what should change now

- We propose that PRO instruments for use in RCTs should be designed with items of low semantic complexity (i.e. semantic molecules), and should be markedly unidimensional, to avoid (or lessen) the RS effect on scores, and enable straightforward interpretation of changes in HRQL scores that might be observed.

INTRODUCTION

The evaluation of the patient perspective in health-related research

Patient-Reported Outcomes (PRO) are now widely used in health-related research, some of them to assess Health-Related Quality of Life (HRQL), usually via self-administered questionnaires [1]. In many medical areas (e.g. oncology, palliative care...), HRQL is measured over time to add relevant information on patients' subjective experience in the course of treatment, to counterbalance objective data such as survival time [2]. Indeed, improvement or deterioration of outcomes like health status or symptom levels is not always correlated with patients' subjective experience [3].

The current generation of HRQL measures is based on the assumption that the meaning of concepts and measurement scales remains stable in individuals' minds over time and is similar between groups [4]. Thus, HRQL scale scores are assumed to be directly comparable for a given individual over time [5]. As illustrated by what were initially called "paradoxical and counter-intuitive findings" from the 1980s and 1990s [6] (e.g. reports of stable HRQL levels over time by patients with a life-threatening disease [7], reports of better levels of HRQL by patients with advanced stages chronic illness than by others [8]...), these assumptions can be challenged. Indeed, these abovementioned findings were interpreted as evidence that respondents understand the same questions differently over time [6,9], a phenomenon which is now known as *response shift (RS)*.

A brief overview of RS theory

In health-related research, RS was defined in 1999 as "a change in the meaning of one's self-evaluation of a target construct" [6]. It is operationalized in three forms:

- *recalibration*, which is a change in the respondent's internal standards of measurement (e.g. a person suffering from chronic pain and rating it on a pain scale as 7/10 will later rate it as 5/10 after experiencing acute pain, despite the chronic pain being the same as before);
- *reprioritization*, which is a change in the respondent's values (i.e. the relative importance of component domains in the target construct - e.g. an athletic person who considers physical functioning as an important part of his/her HRQL may later place emphasis on social functioning after sustaining permanent physical injury);

- *reconceptualization*, which is the redefinition of a target construct (e.g. an item of a multidomain questionnaire initially assessing the domain of mental health, will be later understood by the respondent as assessing another domain, like social functioning) [6].

RS effect is assumed to result from psychological mechanisms that individuals use to deal with life changes, triggered by change in health state (a “catalyst”) [6,10]. As illustrated by [Figure 1](#), when someone is affected by a catalyst (e.g. occurrence of a chronic disease, initiation of chemotherapy...), this catalyst can have a direct effect on HRQL (S_1 pathway in [Figure 1](#)), translating into a change in the person's HRQL assessment. A person's background (e.g. socioeconomic status, personality traits...) can also have a direct effect on HRQL (S_2 pathway) or an effect mediated by the catalyst (S_3 pathway). These effects can be called “standard influences” on HRQL. However, the catalyst can also induce psychological mechanisms (e.g. coping strategies, social comparison...), leading into changes in the way that a person understands and appraises HRQL (R_1 and R_2 pathways), and hence affecting his/her observed scores: RS has occurred.

Thus, there is a need to disentangle RS effect from the effects of the “standard influences” on HRQL [11]. Various methods have been used to detect RS [12–21], although there was in the last years a focus on methods based on Structural Equation Modeling. Indeed, some recent works were helpful in showing how response shift effect can affect observed scores or true attributes (i.e. HRQL itself) and how this effect can be modeled in explaining changes in observed scores or HRQL over time [15,22]. Occurrence of RS has now been documented in a variety of medical conditions [23].

The particular context of clinical trials

HRQL measurement is increasingly used in the context of Randomized Control Trials (RCTs), as endpoints [24]. In this context, a questionnaire should be designed to allow clear interpretation of the direct effect of treatment on patients’ subjective experience. Indeed, it serves as a criterion to enable a decision to be reached between two mutually-exclusive options (i.e. the treatment assessed is effective or not) [25].

The initiation of treatment in the different arms of a RCT can be the catalyst of RS effect. For example, experiences of extreme levels of fatigue after chemotherapy can induce recalibration RS when assessing fatigue [26]. Therefore, if RS has occurred, changes in HRQL scores observed over time cannot be solely explained by a direct effect of the treatment. Moreover, the quality and quantity of RS effect are dependent of the nature of the

catalyst (e.g. as an extreme case, RS is not likely to occur when using placebo in an open label study). Thus, when assessing different treatment options in a RCT, RS effect can affect in varying amounts changes in scores in each arms of the trial. As RS can be triggered by the initiation of treatment (after randomization), it might not be randomly distributed between groups and cannot be equated with measurement error.

Thus, RS effect in a RCT is liable to confuse interpretation of changes in HRQL scores observed over time, thus making conclusions difficult.

HYPOTHESIS

This paper is positioned in the context of the need to measure patients' subjective experience as an endpoint in RCTs. A well-designed questionnaire for use in this context should generate a score that enables simple and clear interpretation of a change over time. It should therefore not be biased by RS, as RS is not usually the effect being investigated.

Our hypothesis is that there is a link between RS effect on HRQL scores (interfering with the effects of "standard influences"), and the level of complexity of HRQL conceptualization (both in semantic and psychometric terms). Indeed, to allow a difference in interpretability over time or between people, the construct being measured has to be defined itself with a sufficient level of complexity.

More precisely, we will discuss below how the impact of reconceptualization on scores depends mostly on the complexity of the definition of a subjective construct (in semantic terms), and how for reprioritization it depends mostly on construct dimensionality (in psychometric terms).

We will then suggest that the impact of RS on the measurement of patients' subjective experience could be avoided (or at least lessened) if subjective constructs and questionnaires were conceptualized and designed with these linguistic and psychometric considerations in mind.

POLYSEMY AS THE ORIGINAL SIN OF HRQL

Linguistics and semantic primes

In the general framework of survey development, the impact of the wording and phrasing of the questionnaire in the response process is discussed. Tourangeau et al. considers that a person who is responding to a survey has to go through four cognitive processes

(comprehension, retrieval, judgment and response) [27]. In this model of survey response, it is stressed out that the first process involved is accurate comprehension of the scope of the questionnaire. Thus, certain characteristics of the items, including semantic ones, are acknowledged as leading to difficulties in understanding precisely what is the subjective construct being measured by the questionnaire [27].

This can be linked to the more specific issue of RS in measuring patients' subjective experience via PRO instruments. Indeed, a PRO instrument can be viewed as a way to communicate from a researcher to a patient [28]. A researcher, through the questionnaire, is aiming to ask patients to rate certain subjective aspects of their lives. RS theory suggests that this communication can be distorted, because the construct being measured can be interpreted in different ways [28]. Although RS has mostly been investigated in HRQL [11], it could occur with any subjective construct that researchers might want to measure (e.g. it has been investigated in relation to fatigue [29]).

This raises the question of whether any subjective construct can be interpreted in different ways by different people (or by one and the same person over time). Or, at least, the extent to which a subjective construct can be interpreted in different ways: i.e. semantics. In the broadest way, semantics is the question of the relations between signifiers and what they stand for. Of course, this question has been studied extensively over the course of human history. It has penetrated many academic fields, like linguistics, philosophy, even cognitive sciences, and it seems that there is currently no definitive consensus about it [30]. However, one of the main semantic theories based on the notion of *semantic primes* and developed since the 1970s by Anna Wierzbicka and colleagues [31], seems to be particularly relevant in the context of this work.

According to Anna Wierzbicka and colleagues, defining every concept used in a language is feasible starting from a core set of primitive concepts: *semantic primes*. Semantic primes are core elements that can be used to coherently represent all complex meanings. They can be viewed as analogous to basic chemical elements from which all chemical compounds can be synthesized [32]. Without a set of primitives, any descriptions of meaning are actually or potentially circular [31].

The search to identify potential semantic primes in languages was based on two main criteria. First, a concept is considered to be a semantic prime if, after extensive trial-and-error lexical-conceptual analysis, the concept cannot be reduced to simpler concepts that define it (reductive paraphrase: it is impossible to define the concept in another way than by paraphrasing it) [31,32]. In addition, a concept can be considered as a semantic prime if, after

extensive cross-cultural studies in a wide range of languages, the semantic prime exists as a linguistic exponent (word or word-like element) in all languages [31,32]. Thus semantic primes represent innate concepts, with a meaning assumed to be universal (i.e. shared by every culture and language) [31].

To date, 64 concepts have been confirmed to be semantic primes (listed in [Table 1](#)) [33].

Looking at [Table 1](#), it seems that none of these semantic primes are strictly correlated with a subjective construct used in health-related research. We cannot therefore postulate that there are subjective constructs relevant to biomedical research that have a universal and univocal meaning. Nevertheless, semantic primes theory describes different levels of complexity in the definition of concepts.

A very low level of complexity, just above the notion of semantic prime, is represented by concepts that can be described using a set of only a few semantic primes. These concepts can be understood as *semantic molecules* [31]. According to Anna Wierzbicka and colleagues, certain concepts related to emotions or feelings can be defined as semantic molecules. As shown in [Figure 2](#), “sadness” can be defined using a few semantic primes combined by means of a universal basic syntax (the “Natural Semantic Metalanguage” [31]).

A higher level of complexity in the definition of concepts is reached when it is at least necessary to define a concept step-by-step, i.e. when first a semantic molecule needs to be formed to define a more complex concept. For instance, to define “face”, first the definition of the semantic molecule “head” is required. “Face” will then be defined using the semantic molecule “head” in its definition [31]. Others levels of complexity are then also described [31].

We hypothesize that there is *a set of subjective constructs* useful in health-related research that could be understood as *semantic molecules*. The complexity of their definition is low; therefore their meaning is easy to convey ([Figure 3](#)). We used “sadness” as an example, but it might be plausible to define the concept of “pain intensity” (we are referring here to the phenomenological experience of pain) as a semantic molecule. The low level of complexity of this definition could be related, in part, to the fact that some subjective constructs can be measured using a single-item instrument ([Figure 3](#)). For instance, the measurement of pain intensity is frequently achieved using a simple Visual Analogue Scale (VAS).

On the other hand, the definition of *other subjective constructs* is of a *higher level of complexity* ([Figure 3](#)). At least, the definition of semantic molecule(s) is first required to describe the concept. In that case, this high level of complexity of definition could be related,

in part, to the need for multiple-item scales to measure a targeted construct. According to current views in the international scientific community, HRQL is highly complex to conceptualize [1,2,34,35]. Prior to the development of the RS theory, there was already some research focusing on the dynamic nature of the HRQL construct [36]. It was then already acknowledged that the meaning of that construct fluctuates across individuals and time [36]. Thus it seems that HRQL is a concept that falls into the category of a subjective construct with a definition entailing a high level of complexity (Figure 3).

To briefly synthesize, we postulate a typology of subjective constructs used in health-related research based on the complexity of their definition:

- there is a set of subjective constructs that can be understood as *semantic molecules* (i.e. their definition exhibits a low level of complexity);
- other subjective constructs exhibit a higher level of complexity (e.g. HRQL).

Relationship between the level of complexity of a construct and reconceptualization

RS

We previously hypothesized that there is a set of subjective constructs useful in health-related research that can be defined as *semantic molecules*. As such, the meaning of these constructs, because it is closely related to the meaning of the few semantic primes used to describe them, if not strictly universal and unambiguous, exhibits a very low level of complexity, and therefore is easy to communicate. If we consider that “pain intensity” referred to earlier is one of these constructs, the frequent use of a single-item scale to measure it could be related to this low level of complexity. An investigator only needs one (or few) item(s) and few instructions to capture most of the relevant information, and to be understood by the respondent.

On the other hand, the meaning of *constructs with a highly complex definition*, such as HRQL, is harder to convey. A multi-item instrument is needed to capture sufficient information and cover all the facets of these complex concepts.

In terms of the operationalization of reconceptualization RS, we postulate that the impact of reconceptualization on a measure is not the same according to the level of complexity of the definition of the constructs explored (Figure 4).

Indeed, for *constructs that can be equated with semantic molecules*, we postulate that they are unlikely to be reconceptualized by given individuals over time, and that if they are presented with the appropriate wording and response scale [27], they will be understood in a very similar way between individuals (Figure 4).

Conversely, *constructs with a highly complex definition*, such as HRQL, are prone to reconceptualization RS, which can affect HRQL scores (Figure 4).

MULTIDIMENSIONALITY AS THE NEXT SIN OF HRQL

The current conceptualization of HRQL

HRQL is also often viewed as highly complex in psychometric terms.

Indeed, although there is no current consensus, whether on the number of domains involved, or on the content of those domains [1,37], most researchers consider HRQL to be a *multidomain* concept - often including physical, social, psychological and spiritual domains - [1,37]. In addition, a broad variety of outcomes, from symptoms and functioning assessment to more subjective appreciations, are frequently subsumed under the umbrella term HRQL, overlapping with the term PRO [25,35].

It is tempting here to make a jump from concept to psychometrics by associating a *multidomain concept* with a *multidimensional construct*. In fact, when researchers are assessing the factorial validity of a PRO instrument using data from a validation sample, it is expected that an instrument designed to measure a multidomain concept will prove to be multidimensional after factor analysis [1,38]. Thus, as HRQL is often conceptualized as a multidomain concept, it is often operationalized to be measured as a *multidimensional construct* (in psychometric terms). One example is the factor structure of the SF-36 Health Survey [39,40].

Relationship between the dimensionality of a construct and reprioritization RS at domain-level

In line with what we previously postulated on the relationship between reconceptualization and the complexity of a construct (in linguistics terms), we now postulate that the impact of reprioritization on measures will vary according to the dimensionality of a construct (in psychometric terms).

As highlighted by Sprangers and Scwhartz in their definition of RS [6], reprioritization implies a change in the relative importance in the respondent's mind between at least two domains constituting the targeted construct. We have also previously seen that a correspondence between domain (in terms of conceptualization) and dimensionality (in terms of operationalization) is often determined by assessing the factor validity of a construct.

Therefore, a subjective construct operationalized as markedly unidimensional may not be a candidate for reprioritization: reprioritization may only occur if a construct is at least bidimensional (Figure 4).

THEORETICAL PROPOSALS TO ACHIEVE A MEANINGFUL SCORE MEASURING PATIENTS' SUBJECTIVE EXPERIENCE FOR CLINICAL TRIALS

To summarize, while there is clearly a need to measure patients' subjective experience in health-related research, in the context of a RCT, a measure of this type needs to be simple and clear to interpret. It should not be biased by RS, as it is usually not the effect of interest. However, HRQL is currently conceptualized in highly complex manner, both in terms of semantic definition (definitions with a high level of complexity according to the semantic primes theory) and in terms of factorial structure (often operationalized as multidimensional), making HRQL scale scores prone to reconceptualization and reprioritization effects.

Thus, we would like to propose that this issue could be efficiently dealt with in the design phase of a PRO instrument, with a rule of "the least linguistic and psychometric complexity" in mind.

Indeed, if a construct leading to a PRO instrument assessing patients' subjective experience of the course of treatment was soundly define (i.e. with the least semantic complexity), then this construct could be broken down into items that can be equated as closely as possible with semantic molecules; and if it was operationalized in unidimensional form, it would be possible to obtain:

- a score that could not be readily biased by RS (reconceptualization and reprioritization);
- a score that could be used as an easily interpretable endpoint in RCTs (because it provides a single score, not a profile [25]).

DISCUSSION

In this paper, we discussed links between complexity of a subjective construct and reconceptualization and reprioritization, but not recalibration RS. Indeed, recalibration is related to the metric standard associated with response categories. In Tourangeau et al. psychological model of survey response, setting the metric standard and mapping a judgment onto response categories are part of the judgment and response processes [27]. Therefore, it

seems that recalibration can be thought as mainly independent of the issue of semantic and psychometric complexity. However, we would like to propose that the problem of recalibration could be dealt with using a suitable response format. Indeed, recalibration can occur because the metric standard is implicit and internal when using a VAS or Likert-scale. We therefore hypothesize that by using brief case vignettes (as in some instruments designed to assess global impression, like the improved Clinical Global Impression Scale (iCGI) [41]), illustrating what each response choice is referring to, the standard metric is rendered explicit and external, forcing the respondent to calibrate himself/herself on the metric that the researcher expects.

In this paper, we hypothesize a link between reprioritization and dimensionality, in agreement with the theoretical definition of reprioritization, which seems to define reprioritization at domain level [6]. We are nevertheless aware that in SEM-based methods, reprioritization is operationalized as a change in values of factor loadings between two measurement times.[14]. Therefore, in SEM operationalization, a change in the values of factor loadings will be viewed as “reprioritization” (it will be “reprioritization at item-level”), even if the construct is unidimensional. It seems that there may be a need here for clarification of the interpretation of changes in parameters of SEM models when performed at item level on a unidimensional construct; or, at least, clarification of the definition of reprioritization.

RS theory can be viewed as related to two different conceptual issues in psychometrics [42]. It is a theory designed to highlight the influence of psychological mechanisms called upon when an individual has to deal with life changes [6]. But as it highlights a potential change in the way that people appraise their HRQL [10], it can also be viewed as a source of bias when it is not the effect investigated by the HRQL measurement [15,28]. Our proposals are related to the latter conceptualization. However, we claim that these two views can be thought to be complementary and context-dependent views: using one or the other depends on the purpose of a study. In the context of a RCT, the direct effect of a treatment on HRQL is usually investigated, so that RS can be viewed as a source of bias [26]. On the other hand, if the purpose is to investigate the different variables affecting a person's HRQL, then it is more appropriate to view RS as one of the outcomes of interest. Therefore, the use of a method to explicitly quantify RS is better suited.

In relation to these last comments, we acknowledge, as has been suggested by some earlier studies [43,44], that psychological interventions can be used to actually induce RS in a patient's mind, as a way to enhance his/her HRQL, especially in the context of palliative care, when curing the disease is no longer an option. If the effectiveness of an intervention of this

type were then to be investigated in the context of a RCT, since RS would be the main effect of interest, a method designed to discriminate the RS effect would be better suited than an instrument designed according to our proposals.

When designing a PRO instrument, establishing validity of the questionnaire is of paramount importance. In the last century, a great deal of efforts has been devoted to successfully develop techniques to investigate structural validity [1]. Although developments have been made via, for instance, the field of “Cognitive Aspects of Survey Methodology”, to reduce the cognitive burden associated with the syntax of items [27,45–48], methodological advances addressing issues regarding face or content validity are still needed to be built. Potential tools derived from semantic primes theory could be developed to help assessing the semantic complexity of construct and items of actual or new HRQL questionnaires. It could also allow for experimental investigation of the theory, by developing items with different semantic complexity and investigate whether it influences the occurrence of RS effect. These potential tools could be complementary options to help researchers to improve face and content validity of a questionnaire.

CONCLUSION

The development of the RS theory has been helpful in highlighting the impact of psychological mechanisms on the assessment of individual HRQL over time. It has also shown that the RS effect could be a source of bias in the context of a RCT, when interpreting an observed change over time in HRQL. Hence our proposal as theoretical guidelines for designing an instrument providing a score that is not biased by RS, with a straightforward interpretation.

We have also highlighted links between RS and HRQL conceptualization and operationalization. Therefore, debate on what HRQL is seems relevant in the wake of the RS theory, in order to clarify what content an investigator wants to convey to a respondent when measuring HRQL using a questionnaire.

We have seen that the “simpler” is the definition of constructs and items, the “easier” is the interpretation of the score measuring patients’ subjective experience of the course of treatment. Semantic primes theory has been helpful to define what is “simple”, in semantic terms. A next step would be to convert it into a tool assessing the level of complexity of the definition of constructs and items when designing PRO instruments.

ACKNOWLEDGMENT

The authors would like to thank Angela Verdier for her help for the English wording and phrasing of the paper.

REFERENCES

- [1] Fayers PM, Machin D. Quality of life: the assessment, analysis, and interpretation of patient-reported outcomes. 2nd ed. Chichester ; Hoboken, NJ: J. Wiley; 2007.
- [2] Leplège A, Hunt S. The problem of quality of life in medicine. *JAMA* 1997;278:47–50.
- [3] Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995;273:59–65.
- [4] Ahmed S, Schwartz C, Ring L, Sprangers MAG. Applications of health-related quality of life for guiding health care: advances in response shift research. *Journal of Clinical Epidemiology* 2009;62:1115–7.
- [5] Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and Quality of Life Outcomes* 2004;2:16.
- [6] Sprangers MAG, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine* 1999;48:1507–15.
- [7] Andrykowski M, Brady M, Hunt J. Positive psychosocial adjustment in potential bone marrow transplant recipients: cancer as a psychosocial transition. *Psychooncology* 1993;2:261–76.
- [8] Cassileth B, Lusk E, Tenaglia A. A psychological comparison of patients with melanoma and other dermatological disorders. *American Academy of Dermatology* 1984;7:742–6.
- [9] McClimans L. A theoretical framework for patient-reported outcome measures. *Theoretical Medicine and Bioethics* 2010;31:225–40.
- [10] Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health Qual Life Outcomes* 2004;2:14.
- [11] Barclay-Goddard R, Epstein JD, Mayo NE. Response shift: a brief overview and proposed research priorities. *Quality of Life Research* 2009;18:335–46.
- [12] Schwartz CE, Sprangers MAG. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science & Medicine* 1999;48:1531–48.
- [13] Schwartz CE, Sprangers MAG. Guidelines for improving the stringency of response shift research using the thentest. *Quality of Life Research* 2010;19:455–64.

- [14] Oort FJ. Using structural equation modeling to detect response shifts and true change. *Quality of Life Research* 2005;14:587–98.
- [15] Oort FJ, Visser MRM, Sprangers MAG. Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology* 2009;62:1126–37.
- [16] Lix LM, Sajobi TT, Sawatzky R, Liu J, Mayo NE, Huang Y, et al. Relative importance measures for reprioritization response shift. *Quality of Life Research* 2012;22:695–703.
- [17] Ahmed S, Mayo N, Scott S, Kuspinar A, Schwartz C. Using latent trajectory analysis of residuals to detect response shift in general health among patients with multiple sclerosis article. *Quality of Life Research* 2011;20:1555–60.
- [18] Mayo NE, Scott SC, Ahmed S. Case management poststroke did not induce response shift: the value of residuals. *Journal of Clinical Epidemiology* 2009;62:1148–56.
- [19] Li Y, Rapkin B. Classification and regression tree uncovered hierarchy of psychosocial determinants underlying quality-of-life response shift in HIV/AIDS. *Journal of Clinical Epidemiology* 2009;62:1138–47.
- [20] Ahmed S, Mayo NE, Wood-Dauphinee S, Hanley JA, Cohen SR. Using the patient generated index to evaluate response shift post-stroke. *Quality of Life Research* 2005;14:2247–57.
- [21] Li Y, Schwartz CE. Data mining for response shift patterns in multiple sclerosis patients using recursive partitioning tree analysis. *Quality of Life Research* 2011;20:1543–53.
- [22] King-Kallimanis BL, Oort FJ, Visser MRM, Sprangers MAG. Structural equation modeling of health-related quality-of-life data illustrates the measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology* 2009;62:1157–64.
- [23] Schwartz CE, Bode R, Repucci N, Becker J, Sprangers MAG, Fayers PM. The clinical significance of adaptation to changing health: a meta-analysis of response shift. *Qual Life Res* 2006;15:1533–50.
- [24] Sanders C, Egger M, Donovan J, Tallon D, Frankel S. Reporting on quality of life in randomised controlled trials: bibliographic study. *BMJ: British Medical Journal* 1998;317:1191.
- [25] McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. *BMC Medicine* 2011;9:86.
- [26] Sprangers MA. Response-shift bias: a challenge to the assessment of patients' quality of life in cancer clinical trials. *Cancer Treat Rev* 1996;22 Suppl A:55–62.
- [27] Tourangeau R, Rips LJ, Rasinski KA. *The psychology of survey response*. Cambridge, U.K.; New York: Cambridge University Press; 2000.
- [28] McClimans L, Bickenbach J, Westerman M, Carlson L, Wasserman D, Schwartz C. Philosophical perspectives on response shift. *Qual Life Res* 2013;22:1871–8.
- [29] Westerman MJ, The A-M, Sprangers MAG, Groen HJM, van der Wal G, Hak T. Small-cell lung cancer patients are just “a little bit” tired: response shift and self-presentation in the measurement of fatigue. *Qual Life Res* 2007;16:853–61.

- [30] Ricœur P. *Hermeneutics and the human sciences: essays on language, action, and interpretation*. Cambridge [Eng.]; New York: Paris: Cambridge University Press; Editions de la Maison des sciences de l'homme; 1981.
- [31] Wierzbicka A. *Semantics: primes and universals*. Oxford [England]; New York: Oxford University Press; 1996.
- [32] Goddard C, Wierzbicka A. *Semantic primes and cultural scripts in language learning and intercultural communication*. *Applied Cultural Linguistics: Implications for second language learning and intercultural communication*, Amsterdam: Gary Palmer and Farzad Sharifian; 2007, p. 105–24.
- [33] Goddard C. *Semantic primes, semantic molecules, semantic templates: Key concepts in the NSM approach to lexical typology*. *Linguistics* 2012;50:711–43. doi:10.1515/ling-2012-0022.
- [34] Brow JP, McGee HM, O'Boyle CA. *Conceptual approaches to the assessment of quality of life*. *Psychology & Health* 1997;12:737–51.
- [35] Leventhal H, Colman S. *Quality of life: A process view*. *Psychology & Health* 1997;12:753–67.
- [36] Allison PJ, Locker D, Feine JS. *Quality of life: a dynamic construct*. *Soc Sci Med* 1997;45:221–30.
- [37] Bakas T, McLennon SM, Carpenter JS, Buelow JM, Otte JL, Hanna KM, et al. *Systematic review of health-related quality of life models*. *Health Qual Life Outcomes* 2012;10:134.
- [38] Brown TA. *Confirmatory factor analysis for applied research*. New York: Guilford Press; 2006.
- [39] Ware JE Jr, Kosinski M, Gandek B, Aaronson NK, Apolone G, Bech P, et al. *The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project*. *International Quality of Life Assessment*. *J Clin Epidemiol* 1998;51:1159–65.
- [40] Keller SD, Ware JE Jr, Bentler PM, Aaronson NK, Alonso J, Apolone G, et al. *Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: results from the IQOLA Project*. *International Quality of Life Assessment*. *J Clin Epidemiol* 1998;51:1179–88.
- [41] Kadouri A, Corruble E, Falissard B. *The improved Clinical Global Impression Scale (iCGI): development and validation in depression*. *BMC Psychiatry* 2007;7:7.
- [42] Fayers PM, Hays RD. *Assessing quality of life in clinical trials: methods and practice*. Oxford; New York: Oxford University Press; 2005.
- [43] Schwartz CE, Sendor RM. *Helping others helps oneself: response shift effects in peer support*. *Social Science & Medicine* 1999;48:1563–75.
- [44] Schwartz CE. *Teaching coping skills enhances quality of life more than peer support: results of a randomized trial with multiple sclerosis patients*. *Health Psychol* 1999;18:211–20.
- [45] McColl E, Meadows K, Barofsky I. *Cognitive aspects of survey methodology and quality of life assessment*. *Qual Life Res* 2003;12:217–8.
- [46] Jobe JB. *Cognitive psychology and self-reports: Models and methods*. *Quality of Life Research* 2003;12:219–27.

[47] Collins D. Pretesting survey instruments: an overview of cognitive methods. *Quality of Life Research* 2003;12:229–38.

[48] Bjorner JB, Ware JE, Kosinski M. The potential synergy between cognitive models and modern psychometric models. *Quality of Life Research* 2003;12:261–74.

ACCEPTED MANUSCRIPT

Table 1. List of semantic primes (English exponents), grouped into related categories, according to Goddard [32]

Figure 1. Theoretical model of relationships between “standard influences”, response shift and changes in HRQL (Adapted from Rapkin and Schwartz [10])

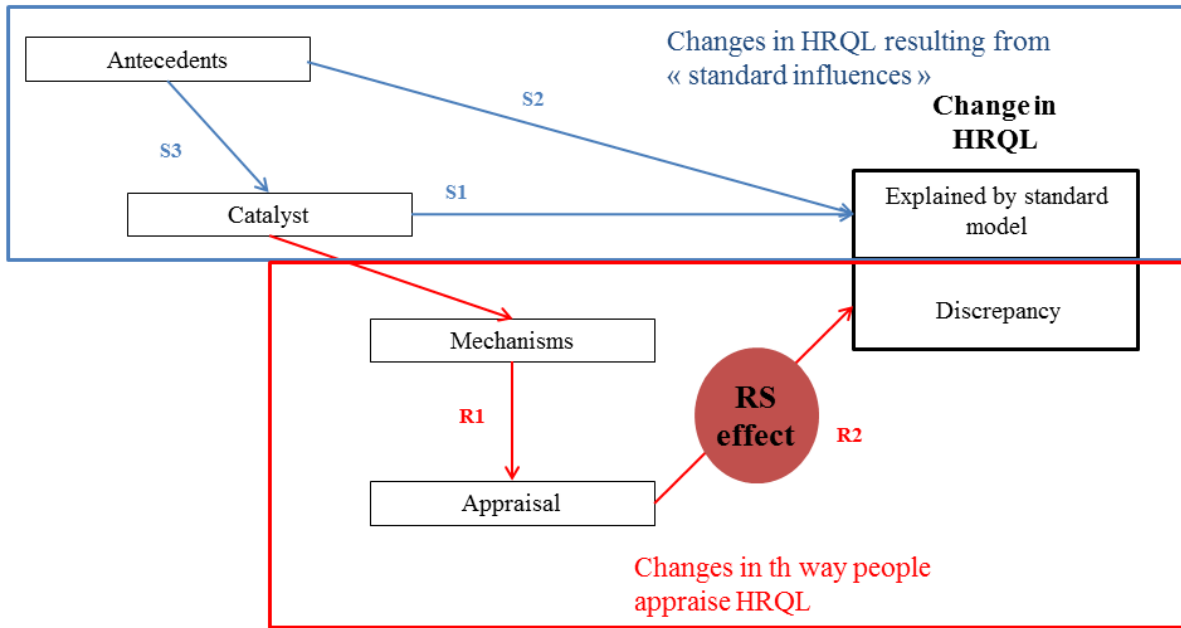
Figure 2. Example of how a few semantic primes are combined to define the semantic molecule "sad", according to semantic primes theory (Adapted from Wierzbicka [30])

Figure 3. Categorization of two sets of subjective constructs useful in health-related research according to the complexity of their linguistic definition

Figure 4. Relationships between complexity of linguistic definition, dimensionality and impact of response shift effects on changes in HRQL scores observed

Table 1. List of semantic primes (English exponents), grouped into related categories, according to Goddard [32]

| Semantic primes | Categories |
|---|--|
| I, You, Someone, Something~Thing, People, Body | substantives |
| Kind, Part | relational substantives |
| This, The same, Other~Else | determiners |
| One, Two, Much~Many, Little~Few, Some, All | quantifiers |
| Good, Bad | evaluators |
| Big, Small | descriptors |
| Know, Think, Want, Feel, See, Hear | mental predicates |
| Say, Words, True | speech |
| Do, Happen, Move, Touch | actions, events, movement, contact |
| Be (Somewhere), There is, Have (Something), Be (Someone/Something) | location, existence, possession, specification |
| Live, Die | life and death |
| When~Time, Now, Before, After, A long time, A short time, For some time, Moment | time |
| Where~Place, Here, Above, Below, Far, Near, Side, Inside | space |
| Not, Maybe, Can, Because, If | logical concepts |
| Very, More | intensifier, augmentor |
| Like~As~Way | similarity |



A Few semantic primes

Something
Feel
Not
Happen Combined to define
Bad
Because
 ...

"Sad (e.g. X feels sad) = a semantic molecule

X feels something

sometimes a person thinks something like this:

something bad happened

if i didn't know that it happened

i would say: i don't want it to happen

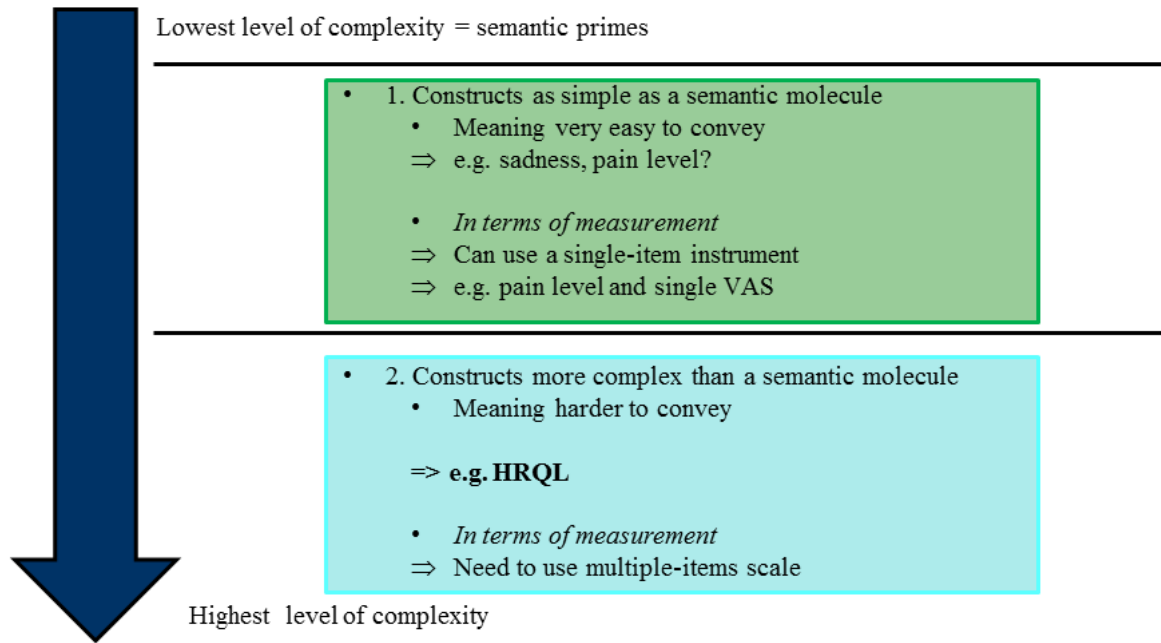
i don't say this now

because i know: i can't do anything

because of this, this person feels something bad

X feels something like this."

ACCEPTED MANUSCRIPT



| | |
|--|--|
| <ul style="list-style-type: none">• 1. Constructs as simple as a semantic molecule<ul style="list-style-type: none">=> Meaning very easy to convey=> Can use a single-item instrument | <ul style="list-style-type: none">• Response shift operationalization<ul style="list-style-type: none">→ Unlikely to be reconceptualized over time and between individuals→ Reprioritization does not seem relevant to consider |
| <ul style="list-style-type: none">• 2. More complex than a semantic molecule<ul style="list-style-type: none">=> Meaning harder to conveyIf conceived as a multidomain construct<ul style="list-style-type: none">=> e.g. HRQL (current conceptualization) | <ul style="list-style-type: none">• Response shift operationalization<ul style="list-style-type: none">→ Prone to reconceptualization→ Operationalized as multidimensional<ul style="list-style-type: none">=> Prone to reprioritization |

ACCEPTED MANUSCRIPT