COMMENTARY

Advancing quality-of-life research by deepening our understanding of response shift: a unifying theory of appraisal

Bruce D. Rapkin[1] · Carolyn E. Schwartz[2,3]

Abstract

Studies of response-shift phenomena in quality-of-life (QOL) research have grown steadily in the more than two decades of research. As a field, we have been calling a lot of different approaches "response shift" over the years, but the only approach that fully embodies the foundational (Sprangers and Schwartz, Soc Sci Med 48(11):1507–1515, 1999) theoretical model is appraisal. According to the definition proposed in this model, response shift is about individual cognitive change. This paper presents the argument that all response-shift research models need to be grounded in an understanding of cognitive appraisal. We present a reasoned argument for why appraisal is fundamental to QOL response-shift research. We describe current measurement and analytic methods for working with appraisal, and how these methods can be integrated into the current response-shift statistical 'tool box.' We propose future research directions on theory, methods, and cross-calibration of group- and individual-level methods. There are currently three tools available in multiple languages for assessing QOL appraisal processes. They have been tested and used to assess response shift in empirical studies of ~7000 people with chronic illness. The study findings illustrate how appraisal theory and methods can facilitate methodological investigations of and to enhance other response-shift detection methods. Future research directions are proposed to enrich QOL theory, response-shift methods, and interpretation of QOL change over time. Appraisal theory and methods are the closest approximation to a response-shift 'gold standard.' They provide the foundation for understanding response shift and point to a unified theory of QOL.

Keywords Response shift · Appraisal · Life events · Theory · Adaptation

In 1999, Sprangers and Schwartz [1, 2] defined response shift as a change in the meaning of a target construct when an individual experiences a health-state change. These changes involved recalibration, reprioritization, and reconceptualization of the construct, and were posited to explain unexpected results in longitudinal quality-of-life (QOL) studies [1, 2]. Studies of response-shift phenomena in QOL research have grown steadily in the more than two decades of

research since these foundational papers. As the field grows, active investigators aim to set the agenda for the next generation of research. As in all disciplines, different people have awareness of different advances. Traditions get established and there is often little impetus to look beyond the familiar. From our vantage point, the only and best way forward is by taking hold of the issue of cognitive appraisal of QOL, and its ramifications for every other aspect of response-shift research. The purpose of this Commentary is to expand on the rationale for appraisal assessment, how it can help to resolve limitations in other response-shift methods, and where the field of QOL research needs to go next in light of these arguments.

As a field, we have been calling a lot of different approaches "response shift" over the years, but the only approach that fully embodies the foundational Sprangers and Schwartz [1, 3] theoretical model is appraisal. If we ever want a coherent field of study, we have to go back to the basics of this model, and that requires assessment

✉ Carolyn E. Schwartz
carolyn.schwartz@deltaquest.org

1 Division of Community Collaboration & Implementation Science, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA

2 DeltaQuest Foundation, Inc., 31 Mitchell Road, Concord, MA, USA

3 Departments of Medicine and Orthopaedic Surgery, Tufts University Medical School, Boston, MA, USA

of how people think about QOL, in other words "cognitive appraisal." According to the definition proposed in this model, *response shift is about individual cognitive change*. Indeed, one cannot re-conceptualize unless one first conceptualizes. It follows that all measures and analytic techniques used in the field should be validated in terms of their ability to explain this change.

For example, one way that people have attempted to identify response shift is by determining whether the factor structure of a measure changes over time, often as a result of a health-related life event or an intervention [4]. Such changes are not unambiguously response shift. There are many ways that differences in factor structure can arise [5]. Unless these factor-structure changes arise because some people have reconceptualized, reprioritized, or recalibrated their ways of evaluating QOL, findings cannot be attributed to response shift. The tendency to interpret group-based methods in terms of some underlying individual cognitive process demonstrates why *linking these group-level statistical methods to appraisal is central to the field*. Individual phenomenology (i.e., appraisal) and group-level methods must be reconciled to unify and advance this field.

Advances in our field depend on getting a better handle on individuals' appraisal. We need to understand how other existing and emergent methods are related to this criterion. Several response-shift detection methods that have emerged in the last decade are group-level statistical analyses. One must ask if it is even possible that response shift is NOT an individual or psychological phenomenon. As such, can group-level methods detect and isolate response shift at an individual level?

Why such a big deal about appraisal?

If we believe that people are reading, comprehending, and responding in meaningful ways to QOL and all other patient-reported outcome (PRO) measures, then something has to be going on in their heads. Understanding how people respond to different items and instructions requires some insight into what they are thinking about and how they arrive at their answers. In our 2004 paper on the psychometric implications of response shift, we had made the distinction that response shift was relevant only to evaluative PROs, where the measurement process involves a judgment using idiosyncratic criteria [6] (Fig. 1). Whereas performance-based measures do not depend on an individual's interpretation, perception- and evaluation-based measures do. Perception-based measures assess matters for which there is actually a correct answer. Evaluation-based measures focus on an individual's personal experience. The inherent subjectivity of evaluative PROs makes them especially valuable, but especially tricky to validate. There is no external reality check. The only way
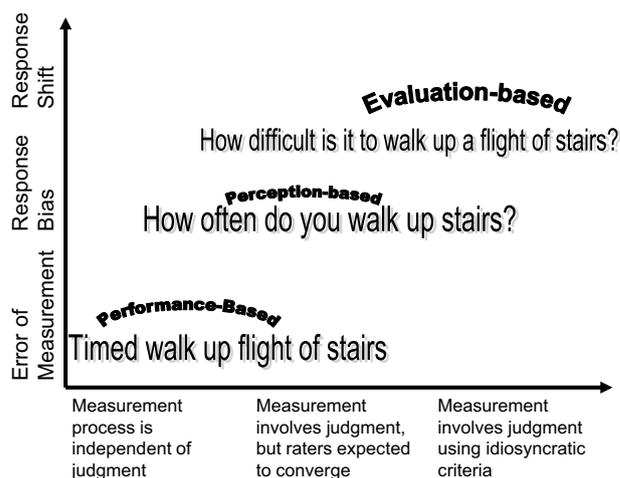
Fig. 1 Clarifying the discrepancy in performance-, perception-, and evaluation-based methods

to understand the basis for an answer is to understand how an individual interprets, evaluates, and frames their response. Capturing change in these subjective appraisals is fundamental to an accurate understanding of change in QOL over time. Characterizing these changes has been a central goal of response-shift research.

Cognitive appraisal of QOL items refers to the direct measurement of the ways that individuals are thinking about QOL questions. Taking into account appraisal helps us to understand why two individuals in identical health states rate the same QOL item differently. It should be emphasized that the fact that two people in exactly the same health state provide different ratings is the entire basis for the field of PRO research, and is built into WHO definitions of health [7]. In other words, the entire justification for our QOL field is that physical health states or observable functioning does not solely determine the psychological experience of health and well-being [8]. Appraisal assessment illuminates the psychological factors that matter in individuals' subjective evaluation of their health and well-being.

There can be no QOL measurement and no PROs without some sort of cognitive appraisal going on. Even if people are automatically reacting to the valence of questions in a detached way with minimal reflection (as may sometimes happen), they can at least reveal that they "weren't thinking about much!" Descriptions of thought processes are challenging and are always filtered through the individuals' meta-cognitions (thinking about how I am thinking). In order to better understand the ways that people think about QOL and PROs in general, our model of QOL appraisal relied on the Sprangers and Schwartz [1] original formulation of response shift as reconceptualization, reprioritization, and recalibration. In order to measure "re-"anything, it implies a change from one time to another. So reconceptualization

requires individual to think about a PRO one way at time one and another way at time two. How they conceptualize at any given time is the appraisal. If the individual demonstrates a change in their conceptualization at a later time, then it is possible that a response shift has occurred. Similarly, people must always have some way to prioritize among the many experiences and moments they might consider in rating a PRO, and must focus on one or another comparator in making ratings. This focus may involve highly intentional reflection or superficial consideration. Whatever the case, it is necessary to determine the thought process of appraisal to understand any given QOL rating at any time. Again, change in how an individual is thinking from one time to the next is a prerequisite to response shift.

One of the advances that drove us toward the use of appraisal was the recognition over 15 years ago that the Sprangers and Schwartz response shift constructs were not easy to operationalize and test [9]. Response shifts generally were inferred from changes in factor structures of QOL outcome measures or the like [10], but there was no direct independent measure of the mechanisms of response shift at the individual level. Response shift was inherent in the outcomes it was intended to explain—in other words, a tautology. The development of the QOL appraisal model [6] remedied that problem (Fig. 2). We posited that three specific conditions had to be met for a response shift to have taken place:

- a change in cognitive appraisal of QOL as described above;
- a discrepancy between the individual's observed QOL rating compared to the rating that the individual would have otherwise made (i.e., unexpected change);
- the ability of change in cognition to explain the unexpected change in QOL, even after taking into account any relevant changes in physical health state, treatments, supports and social determinants.

This model allowed us to link changes in appraisal to changes in QOL ratings without introducing a tautology, because all of these terms could be operationally defined and measured. In this formulation, the Sprangers and Schwartz response shift constructs are epiphenomena that describe specific types of relationships evident in analysis using cognitive factors to explain QOL change. This important conceptual step of identifying response shift as an epiphenomenon underlies many of the methodological developments of the past decade-and-a-half (e.g., Oort SEM [11, 12]; Mayo's Residual Modeling [13]; Lix and Sajobi Relative Importance [14]; Li and Schwartz Classification and Regression Trees and Boucekine Random Forest [16]). These methods *infer* response shift by the behavior of other measureable variables in the model. In contrast, the Rapkin and Schwartz [6] approach *directly assesses appraisal* and defines response shift as the variance explained by changes in QOL that cannot be accounted for by standard predictors of QOL change.

To summarize, appraisal is a big deal because it is fundamental to the experience and the measurement of QOL. If there is not cognition going on when people answer QOL measures, then we have to ask what else is happening when people respond to our measures? Every theoretical formulation or definition of QOL involves subjective contrasts between individuals' understanding of their present state and their desired state. There can be no reconceptualization or reprioritization without conceptualization or prioritization.

Response shift detection methods: then and now

The field of response shift has witnessed notable growth in detection methods. Early work by educational or management science researchers relied on a retrospective-pretest design method called the Then-Test [17, 18]. The method asked respondents to answer one or more PRO questions at posttest with a retrospective appraisal of their baseline status. Recalibration response shift was putatively calculated by simple subtraction of the then-test minus the pretest. Appealing because of its simplicity, the method had been decreasingly used in recent years [10] as evidence built against its validity. It is confounded by recall bias [19, 20] and has
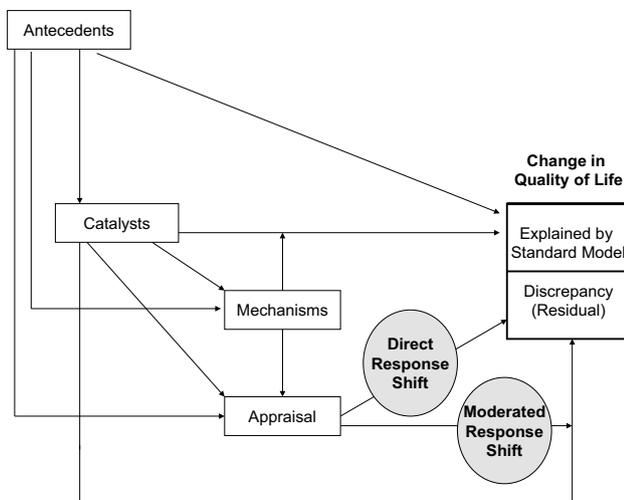
Fig. 2 The Rapkin and Schwartz theoretical model. This model tests for response-shift effects by predicting residualized QOL change. Antecedents are adjusted as covariates in the "standard QOL model", and the unexplained variance (residuals) is modeled as a function of catalysts and appraisal processes. Main effects of appraisal reflect direct response shift, and catalyst-by-appraisal interaction effects reflect moderated response shift

low specificity: qualitative work examining respondent's explanations of their then-minus-pre discrepancy scores revealed that only 15% of responses reflected recalibration [21]. Accordingly, the then-test is a 'noisy' indicator of recalibration. It is also difficult to interpret, with Results and Discussion sections of findings sounding much alike, instead of the latter elucidating the former.

Although its use has continued by early adopters who prefer its simple math and low demand to more current more complex statistical methods, there are numerous options for investigating response shift using recent methodological innovations. For example, Oort's application of SEM methods to quantifying response-shift effects [11, 12] has led to numerous studies of response shift in cancer [9, 22], spinal cord injury [23], primary care [24], and neurological disease [3], to name a few. Other group-level methods promoted in the past decade include the Lix and Sajobi Relative Importance method [14]; the Mayo Trajectory Analysis of Residuals [13]; the ROSALI adaptation of Item Response Theory application building on the Oort SEM approach [25]; Classification and Regression Trees [15, 26] and Random Forest modeling [16]; and Latent Class Analysis [27]. As noted above, all of these statistical advances treat response-shift phenomena as an 'epiphenomenon' that is inferred by the behavior of other variables in the model. International collaborations have produced useful guidelines for secondary analyses investigating response shift [28] as well as cautionary guidelines for working with the then-test [29].

Analytic methods building on the appraisal model

The appraisal model lends itself to a variation of mixed-effects regression called *residual modeling* [30]. Table 1 outlines the analytic steps for this operationalization of the Rapkin and Schwartz [6] model. According to this model, one would include as independent variables indicators for a response-shift catalyst (e.g., health-state change, job loss, marital status change, etc.), antecedents (i.e., stable characteristics of the individual not already included in the 'standard model,' such as personality or expectations), and appraisal change scores. These changes in appraisal may or may not result from intentional adaptive mechanisms (e.g., coping, goal reordering, social comparison). Direct response shift is reflected by main effects of the appraisal change on outcomes, whereas moderated response shifts are reflected by significant interaction effects of appraisal change with catalysts. While one could certainly use the then-minus-pre score as a discrepancy score, we have found that the above residual modeling works well, is easily interpreted, and is an efficient use of repeated-measures data.

How do we measure appraisal?

Three appraisal measures have been developed specifically for use in response-shift research (Table 2). The *QOL Appraisal Profile* (QOLAP) [6] was developed as a first step toward operationalizing the Rapkin and Schwartz theoretical model of appraisal and response-shift [6]. Including both

Table 1 Residual modeling to test response-shift hypotheses

Purpose	Step	Specifics
1. Create residualized change score to operationalize the discrepancy between expected and observed changes	Compute a 'standard QOL model' for each relevant PRO measure	*Statistical method*: random effects models, specify beta coefficients in output for ease of interpretation *Data set structure*: long data set (1 row for each time point within subject) *Dependent variable(s)*: QOL change scores (e.g., PROMIS-10 physical functioning scores follow-up minus baseline for all points of follow-up) *Independent variables*: all the demographic and medical characteristics generally considered important to QOL
	Save the residuals	Residuals are the difference between the observed value of the dependent variable (y) and the predicted value (ŷ)
2. Test models explaining variance in the residuals for each relevant QOL outcome	Test for direct and moderated response-shift effects, after adjusting for antecedents	*Independent variables*: indicators for a response-shift catalyst, antecedents, and appraisal change scores
3. Detect response-shift effects	Interpreting results	*Direct response-shift effects*: significant main effects of appraisal change scores *Moderated response-shift effects*: significant interaction effects for appraisal × catalyst

Table 2 The next-generation QOL appraisal measures

	Original QOLAP	QOLAP v2	Brief Appraisal Inventory
Item format	Open- and close-ended	Close-ended Likert	Close-ended Likert
Number of items	37	85	23
Domains	Appraisal processes • Frame of reference • Standards of comparison • Sampling of experience • Combinatory algorithm	Appraisal processes	Content Pattern of appraisal across parameters
Coding required	Yes	No	No
Analysis unit	Items/codes within domains ($n = 85$ total)	2nd order factor scores ($n = 12$)	1st order factor scores ($n = 5$)
Process of content	Both	Both	Content

open-text and close-ended questions, the tool assesses four appraisal parameters deemed relevant to patient self-report on QOL: frame of reference, sampling of experience, standards of comparison, and combinatory algorithm. In the QOL Appraisal model, the three R's of response-shift are operationalized as follows: Reconceptualization is change in frame of reference; Recalibration is change in standards of comparison; and Reprioritization is change in sampling of experience and/or combinatory algorithm. Over the following years, the validity of the measure was established on a range of patient populations [31–33], and cognitive interviewing studies have documented that patients utilize these four types of appraisal processes when answering QOL questionnaires [34].

Although appraisal is an idiosyncratic process, it is not necessary to measure it idiographically. The recent development of more practical, nomothetic measures of appraisal [35, 36] will facilitate more extensive studies and collaborations. QOLAP data from diverse patient samples were used to identify the 'essence' of appraisal [37] and key goal-delineation domains (i.e., used to derive frame of reference) [38]. Based on these studies, the QOLAPv2 was created. This 85-item close-ended tool has been validated [35] and used in studies of resilience [39], and of the impact of life events on appraisal changes over time [40]. To make it feasible for more researchers to work with appraisal constructs in their clinical research, the *Brief Appraisal Inventory* (BAI) was developed. This 23-item close-ended tool has been validated [36] and used in a study of treatment burden [41]. Both measures have undergone standard protocols for translation [42, 43] into French, Spanish, Italian, German, and Swedish; Mandarin Chinese translations are in progress.

Where does the field need to go next?

There is ample evidence to support our contention that cognitive appraisal processes have to be part of any theory that links changes in health states to individuals' evaluation

of their subjective experience [6, 33, 37, 39, 41, 44–46]. Direct measurement of appraisal represents a breakthrough in understanding QOL because it enables making people's subjective criteria for answering QOL questions explicit. Positing appraisal as the basic mechanism underlying all measurement of PROs raises a wide range of new questions, pointing to future directions in QOL research. These are summarized below:

What psychological constructs are necessary to adequately describe appraisal processes?

The Rapkin and Schwartz QOL Appraisal model [6] used the Sprangers and Schwartz model [1] as its point of departure, and includes measures of conceptualization (frame of reference involving QOL definition and goal delineation), prioritization (sampling of experiences), and calibration (standards of comparison). The model posits that any evaluative QOL item causes people to think about experiences that seem relevant and judge them against salient comparators. People arrive at a single numerical rating by psychologically weighing, balancing, and reconciling different experiences that come to mind (combinatory algorithm). This four-parameter model of appraisal (i.e., frame of reference, sampling of experience, standards of comparison, combinatory algorithm) has proven useful in helping to understand individual differences in response to health-state changes and other life events. However, this does not mean that this particular formulation of appraisal is the final word on the matter. Some questions that require further research include:

a. The four-parameter model implies a process that is linear and orderly. This may not be the most valid representation of how people think in every situation, or at all. People who are more articulate or more introspective likely differ in their meta-cognitive processes. Although almost all people can answer questions about appraisal ("what were you thinking about that led you to rate that

as a 3…?"), it does not mean that the same underlying process occurred. Perhaps some people begin with the visceral or emotional experience ("I feel lousy") and then backfill the experiences and comparators to justify their responses. Research is needed to fully articulate the different ways that people think when answering QOL items.

b. Is QOL appraisal at a given point in time consistent from item to item or fluid? For example, if presented with a set of items about different activities of daily living (ADL), do people recall different experiences or use different comparators, or are their appraisals more coherent? Are people reacting to the content of each item they encounter or responding from a consistent internal "story"?

c. How are processes of appraisal in response to a given QOL question affected by mood or present state? For example, does current fatigue, family concerns or distractions from work lead people to consider different experiences or apply different standards of comparison? Would we get better quality and more consistent data if guided relaxation or mindfulness meditation were used in advance of QOL measurement (like relaxing five minutes before taking a blood pressure measurement)?

d. Appraisal processes in QOL outcomes assessment must be understood in terms of broader cognitive, social, and affective processing of health. For example, how people appraise QOL may be related to how people self-monitor their health status, their ability to self-regulate health behavior, or their sense of control their health outcomes.

e. This formulation of QOL Appraisal also suggests that memory processing and recall play a major role. How is appraisal affected by encoding of memory, state-dependent recall, automatic responding, or trauma? How do cognitive effects of illness and treatment influence appraisal of QOL items?

How should the "expected" QOL score be handled in response-shift research?

As noted earlier, the whole point of response-shift research is to understand why people respond differently to similar changes in health state. This implies some discrepancy between expected and observed scores. Where the "expected" QOL value comes from is an issue that has not been well addressed by our field. In much of the research on response shift, expected values have been determined from sample trends using regression models (e.g., this person's QOL improved much more than we would have expected based on prediction within the sample). However, this sample-derived expected value is not the only way to go. Using changes in appraisal to account for discrepancies relative to observed performance may be useful in some scenarios

(e.g., this person's ratings of ADL-difficulty improved even though their physical fitness or stamina observably declined). The retrospective pretest or "Then-Test" creates a self-defined discrepancy, by asking individuals to indicate what their rating would have been back then using their criteria for appraisal today. Discrepancies could be based on caregiver or provider judgments (this individual rates their improvement as much worse than their provider). It could be particularly useful to understand the criteria that collaterals or providers use to appraise their patient's QOL, in contrast to the patient's own criteria.

How do inferences of response shift derived from group-level analyses relate to response shift identified at the individual level?

Early in the history of response-shift research, there was considerable use of changes in the factor structure of QOL items to infer response shift. For example, if physical and mental health QOL items loaded on the same factor at initial diagnosis with a chronic condition but were on separate, orthogonal factors at a later date, it suggests that the people in the sample were no longer exclusively basing their mental health on their physical condition. Response shift is always and only an individual-level phenomenon. Thus, it is necessary to validate interpretations of response shift based on factor-structure changes in terms of appraisal. For example, the decoupling of mental health from physical health based on change in factor structure over time could only represent response shift if some proportion of the sample stopped basing their ratings of mental well-being on worries related to physical health and started to rely on criteria (for example, social experiences or religious meaning). If observed changes in factor structure cannot be explained by concomitant changes in appraisal, it implies that some other influences may be at play. For example, if some people physically improved but remained worried about their diagnosis, the same change in factor structure could occur even though no individual changed his or her criteria for evaluating either physical or mental health. Of course, validation needs to be a two-way street. If group-level analyses suggest that response shift has occurred but cannot be explained using available measures of appraisal, it is necessary to consider the need for more sensitive cognitive measures. This is especially the case if other explanations are not supported by the data.

Should outcomes studies control for response shift as a bias?

Prior to formulation of the Sprangers and Schwartz model [40] and to the Rapkin and Schwartz [6] formulation of appraisal, response shift was generally understood as a "bias" that needed to be removed to estimate

"true" change. Although Oort's SEM still considers response shift to be a special case of measurement bias, the response-shift effects are treated as useful information rather than error. Treating response-shift effects as bias (i.e., random error to be reduced or removed) and calling adjusted change "true change" are inherently problematic, because this terminology maligns individuals' actual ratings. Was the person answering more accurately or honestly at T1 or T2? Was their way of thinking wrong?

Although it would still be possible to control for changes in appraisal to examine change in QOL scores, it may be more worthwhile to think of QOL as a construct that is subject to several different kinds of change—in evaluative level (traditional rating scales), in breadth of experiences considered, in higher versus lower standards of comparison, and the like. Outcomes could be evaluated in terms of changes across these dimensions. For example, we might want to see that people undergoing treatment improve in their evaluative QOL rating while maintaining high self-expectations about levels of activities or standards of comparison. Alternatively, in some instances changing patients' expectations regarding standards for QOL may be a better outcome than inflated QOL ratings based on denial. Framing hypotheses about PROs in terms of patients' status across multiple psychological dimensions of appraisal is likely to better reflect the clinical reality concerning the kinds of decisions and trade-offs made by clinicians and providers. Future research needs to examine the pros and cons of thinking about QOL outcomes in terms of multiple parameters of appraisal.

Conclusions

In summary, response-shift research has enriched the field of QOL research by reflecting deeply about how individuals' adaptation trajectories impact measured treatment outcomes using evaluative PROs. The work done on appraisal—theory, methods, and empirical research—has helped to characterize how thinking impacts function and outcome, leading to clear clinical applications for improving treatment outcomes in diverse patient populations. We believe that appraisal theory and methods can provide a useful unifying foundation for the field of QOL research in general and response-shift research in particular. We hope that the above suggested future directions will facilitate further growth and understanding of what patients can do to have an impact on the course of their disease.

Compliance with ethical standards

References

1. Sprangers, M. A. G., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: A theoretical model. *Social Science and Medicine, 48*(11), 1507–1515.
2. Schwartz, C. E., & Sprangers, M. A. G. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science and Medicine, 48*(11), 1531–1548.
3. King-Kallimanis, B. L., Oort, F. J., Nolte, S., Schwartz, C. E., & Sprangers, M. A. G. (2011). Using structural equation modeling to detect response shift in disability and QOL scores of multiple sclerosis patients. *Quality of Life Research, 20*(10), 1527–1540.
4. Ahmed, S., Mayo, N. E., Corbiere, M., Wood-Dauphinee, S., Hanley, J., & Cohen, R. (2005). Change in quality of life in people with stroke over time: True change or response shift? *Quality of Life Research, 14,* 611–627.
5. Edelen, M. O., Thissen, D., Teresi, J. A., Kleinman, M., & Ocepek-Welikson, K. (2006). Identification of differential item functioning using item response theory and the likelihood-based model comparison approach: Application to the Mini-Mental State Examination. *Medical Care, 44,* S134–S142.
6. Rapkin, B. D., & Schwartz, C. E. (2004). Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health and Quality of Life Outcomes, 2*(1), 14.
7. World Health Organization. (1948). World health organization constitution. *Basic documents* (Vol. 1). Geneva: World Health Organization.
8. Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life: A conceptual model of patient outcomes. *JAMA, 273*(1), 59–65.
9. Visser, M. R., Oort, F. J., van Lanschot, J. J. B., van der Velden, J., Kloek, J. J., Gouma, D. J., et al. (2013). The role of recalibration response shift in explaining bodily pain in cancer patients undergoing invasive surgery: An empirical investigation of the Sprangers and Schwartz model. *Psycho-Oncology, 22*(3), 515–522.
10. Sajobi, T. T., Brahmbatt, R., Lix, L. M., Zumbo, B. D., & Sawatzky, R. (2018). Scoping review of response shift methods: Current reporting practices and recommendations. *Quality of Life Research, 27*(5), 1133–1146.
11. Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research, 14,* 587–598.
12. Oort, F. J., Visser, M. R. M., & Sprangers, M. A. G. (2005). An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. *Quality of Life Research, 14,* 599–609.
13. Mayo, N., Scott, C., & Ahmed, S. (2009). Case management post-stroke did not induce response shift: The value of residuals. *Journal of Clinical Epidemiology, 62,* 1148–1156.

14. Lix, L. M., Sajobi, T. T., Sawatzky, R., Liu, J., Mayo, N. E., Huang, Y., et al. (2013). Relative importance measures for reprioritization response shift. *Quality of Life Research, 22*(4), 695–703. https://doi.org/10.1007/s11136-012-0198-3.

15. Li, Y., & Schwartz, C. E. (2011). Using classification and regression tree to examine evidence of quality of life response shift in patients with multiple sclerosis. *Quality of Life Research, 20*(10), 1543–1553. https://doi.org/10.1007/s11136-011-0004-7.

16. Boucekine, M., Loundou, A., Baumstarck, K., Minaya-Flores, P., Pelletier, J., Ghattas, B., et al. (2013). Using the random forest method to detect a response shift in the quality of life of multiple sclerosis patients: A cohort study. *BMC Medical Research Methodology, 13,* 20. https://doi.org/10.1186/1471-2288-13-20.

17. Howard, G. S., Dailey, P. R., & Gulanick, N. A. (1979). The feasibility of informed pre-tests in attenuating response-shift bias. *Applied Psychological Measurement, 3,* 481–494.

18. Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: types of change generated by OD designs. *The Journal of Applied Behavioral Science, 12,* 133–157.

19. Schwartz, C. E., Sprangers, M. A. G., Carey, A., & Reed, G. (2004). Exploring response shift in longitudinal data. *Psychology and Health, 19*(1), 51–69.

20. Ahmed, S., Mayo, N. E., Wood-Dauphinee, S., & Hanley, J. (2001). Response shift in the assessment of health-related quality of life (HRQL) post-stroke. *Quality of Life Research, 10,* 204.

21. Schwartz, C. E., & Rapkin, B. A. (2012). Understanding appraisal processes underlying the thentest: A mixed methods investigation. *Quality of Life Research, 21*(3), 381–388.

22. Verdam, M. G. E., Oort, F. J., van der Linden, Y. M., & Sprangers, M. A. G. (2015). Taking into account the impact of attrition on the assessment of response shift and true change: A multigroup structural equation modeling approach. *Quality of Life Research, 24,* 541–551. https://doi.org/10.1007/s11136-014-0829-y.

23. Schwartz, C. E., Stucky, B. D., Rivers, C. S., Noonan, V. K., & Finkelstein, J. A. (2018). Quality of life and adaptation in people with spinal cord injury: Response shift effects five-years post-injury. *Archives of Physical Medicine and Rehabilitation, 99,* 1599–1608.

24. Ahmed, S., Mayo, N. E., Wood-Dauphinee, S., Hanley, J. A., & Cohen, S. R. (2005). The structural equation modeling technique did not show a response shift, contrary to the results of the then test and the individualized approaches. *Journal of Clinical Epidemiology, 58*(11), 1125–1133.

25. Guilleux, A., Blanchin, M., Vanier, A., Guillemin, F., Falissard, B., Schwartz, C. E., et al. (2015). Response shift algorithm in item response theory (ROSALI) for response shift detection with missing data in patient-reported outcomes in longitudinal clinical trials. *Quality of Life Research, 24*(3), 553–564.

26. Li, Y., & Rapkin, B. D. (2009). Classification and regression tree analysis to identify complex cognitive paths underlying quality of life response shifts: A study of individuals living with HIV/AIDS. *Journal of Clinical Epidemiology, 62,* 1138–1147.

27. Sawatzky, R., Gadermann, A. M., Ratner, P. A., Zumbo, B., & Lix, L. (2012). Identifying individuals with inflammatory bowel disease who experienced response shift: A latent class analysis. *Quality of Life Research, 21,* 33.

28. Schwartz, C. E., Ahmed, S., Sawatsky, R., Sajobi, T., Mayo, N., Finkelstein, J. A., et al. (2013). Guidelines for secondary analysis in search of response shift. *Quality of Life Research, 22,* 2663–2673. https://doi.org/10.1007/s11136-013-0402-0.

29. Schwartz, C. E., & Sprangers, M. A. G. (2010). Guidelines for improving the stringency of response shift research using the Thentest. *Quality of Life Research, 19,* 455–464.

30. Reed, B. R., Mungas, D., Farias, S. T., Harvey, D., Beckett, L., Widaman, K., et al. (2010). Measuring cognitive reserve based on the decomposition of episodic memory variance. *Brain, 133*(Pt 8), 2196–2209. https://doi.org/10.1093/brain/awq154.

31. Rapkin, B. D., & Fischer, K. (1992). Framing the construct of life satisfaction in terms of older adults' personal goals. *Psychology and Aging, 7*(1), 138–149.

32. Rapkin, B. D., & Fischer, K. (1992). Personal goals of older adults: Issues in assessment and prediction. *Psychology and Aging, 7*(1), 127–137.

33. Schwartz, C. E., & Rapkin, B. D. (2012). Understanding appraisal processes underlying the Thentest: A mixed methods investigation. *Quality of Life Research, 21*(3), 381–388. https://doi.org/10.1007/s11136-011-0023-4.

34. Bloem, E. F., van Zuuren, F. J., Koneman, M. A., Rapkin, B. D., Visser, M. R. M., Koning, C. C. E., et al. (2008). Clarifying quality of life assessment: Do theoretical models capture the underlying cognitive processes? *Quality of Life Research, 17,* 1093–1102.

35. Rapkin, B. D., Garcia, I., Michael, W., Zhang, J., & Schwartz, C. E. (2017). Distinguishing appraisal and personality influences on quality of life in chronic illness: Introducing the Quality-of-Life Appraisal Profile version 2. *Quality of Life Research, 26,* 2815–2829. https://doi.org/10.1007/s11136-017-1600-y.

36. Rapkin, B. D., Garcia, I., Michael, W., Zhang, J., & Schwartz, C. E. (2017). Development of a practical outcome measure to account for individual differences in quality-of-life appraisal: The brief appraisal inventory. *Quality of Life Research.* https://doi.org/10.1007/s11136-017-1722-2.

37. Rapkin, B. D., & Schwartz, C. E. (2016). Distilling the essence of appraisal: A mixed methods study of people with multiple sclerosis. *Quality of Life Research, 25*(4), 793–805.

38. Schwartz, C. E., Li, J., & Rapkin, B. D. (2016). Refining a web-based goal assessment interview: Item reduction based on reliability and predictive validity. *Quality of Life Research, 25*(9), 2201–2212. https://doi.org/10.1007/s11136-016-1258-x.

39. Schwartz, C. E., Michael, W., & Rapkin, B. D. (2017). Resilience to health challenges is related to different ways of thinking: Mediators of quality of life in a heterogeneous rare-disease cohort. *Quality of Life Research, 26,* 3075–3088. https://doi.org/10.1007/s11136-017-1633-2.

40. Schwartz, C. E., Zhang, J., Stucky, B. D., Michael, W., & Rapkin, B. D. (2019). The implacability of item-response-theory measures: Does appraisal still impact interpretation of change? *Quality of Life Research, 27*(Suppl 1), S32.

41. Schwartz, C. E., Zhang, J., Michael, W., Eton, D. T., & Rapkin, B. D. (2018). *Reserve-building activities attenuate treatment burden in chronic illness: The mediating role of appraisal and social support* (pp. 1–11). January-June: Health Psychology Open.

42. Cull, A., Sprangers, M., Bjordal, K., Aaronson, N., West, K., & Bottomley, A. (2002). *EORTC quality of life group translation procedure.* Brussels: EORTC.

43. Bullinger, M., Alonso, J., Apolone, G., Leplège, A., Sullivan, M., Wood-Dauphinee, S., et al. (1998). Translating health status questionnaires and evaluating their quality: The IQOLA project approach. *Journal of Clinical Epidemiology, 51*(11), 913–923.

44. Schwartz, C. E., & Rapkin, B. D. (2004). Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and Quality of Life Outcomes, 2,* 16.

45. Schwartz, C. E., Quaranto, B. R., Rapkin, B. D., Healy, B. C., Vollmer, T., & Sprangers, M. A. G. (2014). Fluctuations in appraisal over time in the context of stable and non-stable health. *Quality of Life Research, 23*(1), 9–19. https://doi.org/10.1007/s11136-013-0471-0.

46. Schwartz, C. E., Powell, V. E., & Rapkin, B. D. (2017). When global rating of change contradicts observed change: Examining appraisal processes underlying paradoxical responses over time. *Quality of Life Research, 26,* 847–857. https://doi.org/10.1007/s11136-016-1414-3.