


RESEARCH

Open Access



# Response shift in health-related quality of life measures in the presence of formative indicators

Silvia Testa<sup>1</sup>, Daniela Di Cuonzo<sup>2,3</sup>, Giuliana Ritorto<sup>4</sup>, Laura Fanchini<sup>4</sup>, Sara Bustreo<sup>4</sup>, Patrizia Racca<sup>4</sup> and Rosalba Rosato<sup>2,3\*</sup> 

## Abstract

**Background:** Response shift (RS) has been defined as a change in the meaning of an individual's self-evaluation that needs to be accounted for when assessing longitudinal changes in health-related quality of life (HRQoL). RS detection through structural equation modeling is accomplished by adopting Oort's procedure based on a measurement model in which the observed variables are defined as reflective indicators of the HRQoL latent variable; that is, the latent variable causes the variation in the reflective indicators. This study aims to propose a procedure that assesses RS when formative indicators are used in measuring HRQoL; in this last case, the latent variable is considered to be a function of some formative indicators. A secondary aim is to compare the new procedure with Oort's procedure to highlight similarities and differences.

**Methods:** The data were retrieved from a consecutive series of 258 patients newly diagnosed with colorectal cancer and undergoing chemotherapy and/or surgery. The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QOL-C30) was administered twice, once before and once six months after treatment. Structural equation modeling was used to evaluate RS and true change with the newly proposed method (in which fatigue and pain were defined as formative indicators) and with Oort's procedure (in which fatigue and pain were defined as reflective indicators).

**Results:** According to the new procedure, there was no measurement bias, and on average, patients' quality of life improved by 3.53 points (on a scale ranging from 0 to 100) at the 6-month follow-up. With Oort's procedure, the loading of the pain indicator was not invariant across the two time points, suggesting the presence of reprioritization, whereas the estimation of true change was very similar to the previous one: 3.87.

**Conclusions:** RS and true change in HRQoL can be evaluated in the presence of formative indicators. Defining a measurement model by formative or reflective indicators can lead to different results.

**Keywords:** Response shift, Formative indicators, Reflective indicators, Health-related quality of life

## Introduction

The longitudinal change in patients' perceived health-related quality of life (HRQoL) is a relevant topic in the evaluation of treatment effects and disease adaptation in individuals with chronic disease conditions [1]. Often, self-report HRQoL questionnaires are administered at two or more time points, and differences in the repeated

\*Correspondence: rosalba.rosato@unito.it

<sup>2</sup> Department of Psychology, University of Turin, Turin, Italy

Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

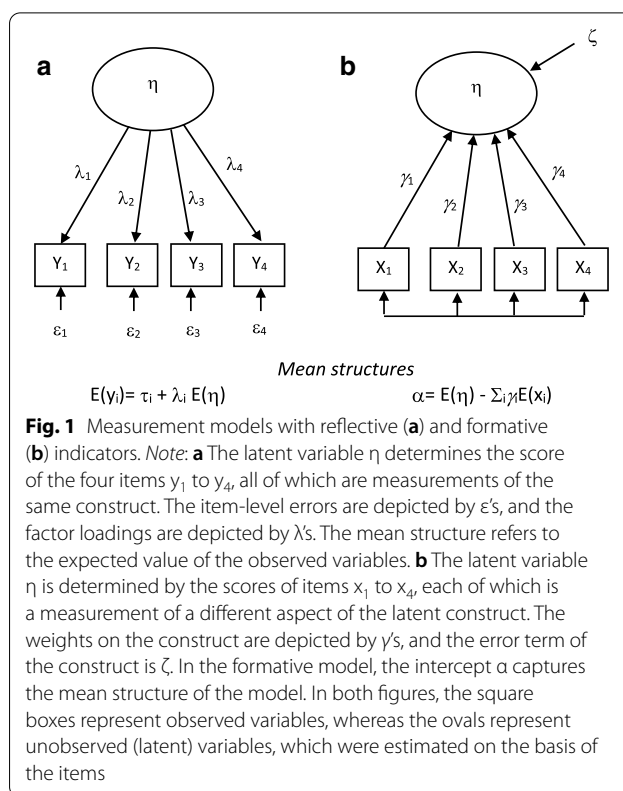
measures are interpreted as changes in the construct of interest, assuming that the meaning of the items has not changed over this interval. However, the validity of this assumption can be undermined by the presence of the so-called response shift (RS) phenomenon [2–5].

As defined by Sprangers and Schwartz in 1999, RS refers to a change in the meaning of one’s self-evaluation of a target construct due to (a) a change in the internal standards of measurement (*recalibration*); (b) a change in the importance of the component domains constituting the target construct (*reprioritization*); or (c) a redefinition of the target construct (*reconceptualization*). Although changes in the meaning of the target construct occur at the individual level, RS can also be analyzed at the group level, assuming that the majority of the cohort shows the same pattern of RS [3, 6, 7]. Moreover, RS can be evaluated at the item level—by analyzing the relationship between the patients’ rating and a narrow latent variable (typically an HRQoL domain, such as a physical, emotional or social quality of life dimension)—and at the domain level by considering the relationship between domain scores (e.g., physical, emotional, social) and the broader latent variable, HRQoL [3]. According to the model proposed in 1999 by Sprangers and Schwartz and subsequent updates [4, 5, 8], RS is triggered by a catalyst (i.e., a relevant event leading to a change in the respondent’s health status) and is related to a change in the appraisal process involved in answering HRQoL items. Thus, RS leads to a discrepancy between the observed change recorded by the patient’s self-evaluations and the true change in the construct of interest.

The structural equation modeling (SEM) approach is often used to model the relationship between the observed scores (item or domain scores) and the target construct (latent variable) [6, 9]; however, there are many studies debating the issue of measurement and reflective/formative models [10–13], and we hereby present a brief description of both types of models.

In a reflective model, constructs are assessed on the basis of the assumption that the underlying latent construct causes the observed variation in the observed variables. The observed measures are called reflective indicators, and they are expected to be correlated because of their dependency on the same latent constructs. For some constructs, it is more appropriate to view causality from the observed indicators to the construct. Such models are called formative, namely, it is assumed that changes in the manifest measures (formative indicators) cause changes in the underlying construct. The path diagrams of the two types of measurement models (reflective and formative) are depicted in Fig. 1.

Formative indicator models such as that shown in Fig. 1b are not identified; to solve the identification issue,



the latent variable needs to affect at least two observed or latent uncorrelated variables [14]. Moreover, formative indicators do not necessarily have to be correlated, as they contribute to the definition of the construct and are not a reflection of it. For example, fatigue and pain are undoubtedly part of the multidimensional definition of HRQoL in the context of health but are not necessarily correlated (patients with a high fatigue score are not expected to have a high pain score, and patients with a low fatigue score do not necessarily have a low pain score; it can happen, but one condition does not necessitate the other). Rather, strong correlations between formative indicators, such as those described by multicollinearity in standard regression models, undermines the stability of the estimates of parameters. Last, formative indicators need to be error-free to avoid biased estimates of  $\beta$ s.

**Detection of RS by using SEM and reflective indicators: Oort’s procedure**

From the several methodological approaches developed to address RS, Oort’s procedure [6] is one of the most attractive [3]. Oort’s procedure applies the reflective model and facilitates the detection of all forms of RS (*recalibration*, *reprioritization*, *reconceptualization*) in longitudinal studies. In Oort’s procedure, the following

model links a set of observed scores  $y_{it}$  ( $i = 1, 2, \dots, S$ ) to the target latent variable ( $\eta_t$ ) at time point  $t$  ( $t = 1, 2$ ):

$$y_{it} = \lambda_{it} \cdot \eta_t + \tau_{it} + \varepsilon_{it} \tag{1}$$

The equation depicts the classic measurement model;  $\lambda_{it}$  is the loading of the indicator  $y_{it}$  on the target construct  $\eta_t$ ,  $\tau_{it}$  is the intercept parameter, and  $\varepsilon_{it}$  is the error term, whose variance is  $V(\varepsilon_{it})$ .

The three forms of RS correspond to specific changes in model parameters ( $\lambda_{it}$ ,  $\tau_{it}$ ,  $V(\varepsilon_{it})$ ) between the two time points. Specifically, when at one—and only one—of the two time points the loading is zero, a *reconceptualization* has occurred; when both loadings are not equal to zero but differ in magnitude, *reprioritization* has occurred. Changes in the internal standards of measurement are signaled by a difference in the  $\tau$  parameters (uniform *recalibration*) and by a change in the error variances,  $V(\varepsilon_{i1})$  and  $V(\varepsilon_{i2})$  (nonuniform *recalibration*).

Oort’s procedure requires items/domains to be defined as reflective (effect) indicators of the target construct, i.e., the observed scores are manifestations of the latent variable [15]. However, some variables of health and quality of life are better conceived as formative (causal) indicators, as they determine the latent variable [16–18].

**Moving from reflective to formative indicators**

In a measurement model with formative indicators, a set of observed scores  $x_i$  ( $i = 1, 2, \dots, M$ ) at time point  $t$  determines the meaning of the target construct  $\eta_t$ :

$$\eta_t = \alpha_t + \gamma_{1t} x_{1t} + \gamma_{2t} x_{2t} + \dots + \gamma_{Mt} x_{Mt} + \zeta_t \tag{2}$$

where  $\alpha_t$  is the model intercept,  $\gamma_{it}$  is a regression coefficient indicating the contribution of each indicator to the formation of the latent variable, and  $\zeta_t$  is the error term, with variance  $V(\zeta_t)$  (Fig. 1).

Residual variance  $V(\varepsilon_i)$  in the reflective-type model is the variance of the indicators not explained by the common latent factor; in the formative-type model,  $V(\zeta)$  is the variance of the latent variables not explained by the formative indicators.

Hence, a formative model involves intercepts ( $\alpha$ s), weights ( $\gamma$ s) and construct residual variance ( $V(\zeta)$ ), which are parameters not included in a reflective model. For this reason, Oort’s procedure cannot be used and, at the best of our knowledge, a specific procedure for this type of model is not available in the literature. In the present study, we aimed to propose a procedure to detect the presence of RS when a measurement model is defined with formative indicators. The secondary objective was to compare the new procedure with Oort’s procedure to highlight similarities and differences.

**Methods**

The procedure to detect RS in the presence of formative indicators was first presented and then applied in a cohort of patients with cancer to whom the European Organisation for Research and Treatment of Cancer (EORTC) quality of life questionnaire (QOL-C30). The instrument was administered at two time points, and the diagnosis was assumed as the catalyst event. Several studies have examined RS among patients with colorectal cancer (CRC) [19–22]. The EORTC QOL-C30 was chosen as the HRQoL questionnaire because previous studies in the literature investigated the nature (reflective vs formative) of its items [23, 24]. Finally, RS and true change were evaluated by the new procedure and compared to those obtained by Oort’s procedure.

**RS detection in the case of a measurement model with formative indicators: a proposal**

The presence of the latent variable error term in the formative-type model is important for differentiating it from a composite measure model in which the composite variable is an exact combination of the indicators (without an error term). As stated by Bollen, “causal indicators tap a unidimensional concept (or dimension of a concept), and the latent variable that represents the concept is not completely determined by the causal indicators” (Bollen [25], p. 360, see also Bollen and Bauldry [26]). In contrast with the composite variable model, the formative-type model deals with latent variables; thus, a definition of RS similar to that proposed by Oort [6] can be applied.

The procedure we present is applicable when RS is assessed at the domain level. We will describe a situation in which two or more reflective indicators are used to overcome the identification issue, and at the end of the paragraph, we will present modifications for a situation in which two or more dependent variables are used for model identification.

The prerequisites are as follows: (a) at least two reflective indicators of the HRQoL construct that have been shown to be invariant across the two time points are available (at least metric and scalar invariant:  $\lambda_{i1} = \lambda_{i2}$  and  $\tau_{i1} = \tau_{i2}$  for all  $i$ ), (b) the formative indicators are reliable and are not strongly correlated, and (c) changes in the mean of each formative indicator between the two time points can be assumed as true changes in the corresponding HRQoL domain (the underlying reflective indicators show at least metric and scalar invariance).

Table 1 reports the operational definitions of *reconceptualization* and *reprioritization*, which are similar to those used in Oort’s procedure. The former is a change in the set of formative indicators that measure

**Table 1 RS detection comparison in changes in SEM parameters between two time points**

Type of change	Reflective (Oort’s procedure)	Formative (Our proposal)
Reconceptualization	Patt. ( $\Lambda_1$ ) $\neq$ Patt. ( $\Lambda_2$ )	Patt. ( $\Gamma_1$ ) $\neq$ Patt. ( $\Gamma_2$ )
Reprioritization	$\lambda_{i1} \neq \lambda_{i2}$	$\gamma_{i1} \neq \gamma_{i2}$
Reconceptualization (secondary)	–	$V(\zeta_1) \neq V(\zeta_2)$
Recalibration (uniform)	$\tau_{i1} \neq \tau_{i2}$	–
Recalibration (nonuniform)	$V(\epsilon_{i1}) \neq V(\epsilon_{i2})$	–
True change in means	$M(\eta_1) \neq M(\eta_2)$	$f(\alpha_1, \alpha_2)$ , where $\alpha_1 = 0$
True change in the construct variance	$V(\eta_1) \neq V(\eta_2)$	–

Patt. means the pattern of zero and nonzero parameters in the matrix of loadings ( $\Lambda$ ) and regression coefficients ( $\Gamma$ ); V and M stand for variance and mean. For the meaning of the other symbols, see the description of Eqs. 1 and 2

the target construct, i.e., one or more regression coefficients that are zero at only one of the two time points; the second is a change in the magnitude of the regression coefficients from one time point to the other. In the presence of a time invariance of  $\gamma$ s, changes in the error variances imply that the variance of the latent variable that is jointly explained by the formative indicators has changed. These changes can be considered indicative of a secondary form of *reconceptualization*: at one of the two time points, specifically at that with the greater variance in error, one or more formative indicators is omitted.

Regarding the mean structure part of the model, another parameter can be tested for time invariance: the model intercept ( $\alpha$ ). When the intercept at  $t = 1$  is constrained to zero, the intercept at  $t = 2$  informs us of the intercept change across the time points, and this change can be used to estimate the true change in the target construct. In our procedure, the model has to be estimated twice, first with mean-centered formative indicators ( $f_{centered}$ ) and second with mean-centered reflective indicators ( $r_{centered}$ ). In the first estimate,  $\alpha_{f\_centered,2}$  reflects the mean change, as detected by the reflective part of the model, whereas in the second estimate,  $\alpha_{r\_centered,2}$  reflects the mean change, as detected by the formative part of the model. The estimate of true change is obtained by computing the mean difference between  $\alpha_{f\_centered,2}$  and  $\alpha_{r\_centered,2}$ , as detailed below.

Starting from the mean structure of the regression equation of a formative indicator model (Fig. 1), the intercept is  $\alpha = E(\eta) - \sum_i \gamma_i E(x_i)$ , and it can be written more compactly as  $\alpha = A - B$ , where  $A = E(\eta)$  and  $B = \sum_i \gamma_i E(x_i)$ .

When  $\alpha_1 = 0$ ,  $\alpha_2$  can be rewritten as follows:

$$\alpha_2 = (A_2 - B_2) - (A_1 - B_1) = (A_2 - A_1) - (B_2 - B_1) \tag{3}$$

If the reflective indicators used to solve the identification issue are mean centered, the intercept is:

$$\alpha_{r\_centered,2} = -(B_2 - B_1) \tag{4}$$

In turn, if the formative indicators are mean centered, the intercept is:

$$\alpha_{f\_centered,2} = (A_2 - A_1) \tag{5}$$

True change can be defined as the average difference between the two intercept values:

$$(\alpha_{f\_centered,2} - \alpha_{r\_centered,2})/2 = [(A_2 - A_1) + (B_2 - B_1)]/2 \tag{6}$$

As can be seen from Eq. 3, if data are not mean centered, the intercept  $\alpha_2$  does not capture the joint contribution of the variation in the reflective indicators, which is reflected in the  $(A_2 - A_1)$  term, and of the variation in the formative indicators, which is reflected in the  $(B_2 - B_1)$  term. Their joint contribution is taken into account by computing the mean of the intercepts estimated after centering the data once with respect to the formative indicators and once with respect to the reflective indicators (Eq. 6).

Let us now consider how to adapt the procedure to the case in which two or more dependent variables (instead of two or more reflective indicators) are used to fix the identification problem. The procedure is the same as before, except regarding the first prerequisite and the way in which true change is evaluated. The first prerequisite states that the two or more dependent variables are mean centered, while the assessment of true change requires a single model estimation, and the true change is directly estimated by the  $\alpha_2$  parameter.

**Study cohort and data-collection procedures**

The study was a prospective study aimed at evaluating changes in quality of life in CRC patients between the time of diagnosis and the six-month follow-up. Two hundred fifty-eight participants were enrolled at the cancer care unit at the Città della Salute e della Scienza Hospital

in Turin between October 2014 and October 2016. The inclusion criteria were a new diagnosis of CRC and an age older than 18 years. Patients with previous neoplasms, cognitive disorders (clinical judgment) or insufficient understanding of the Italian language were excluded. In this prospective study, patients were enrolled at diagnosis during the first multidisciplinary visit to determine the appropriate chemotherapy treatment and were re-evaluated at the six-month follow-up visit (between April 2015 and April 2017). At baseline, the respondents completed forms pertaining to demographic information and their self-reported health status and mood disorder questionnaires; the self-reported questionnaires were readministered at the follow-up visit. A member of the research team was available to help fill out the questionnaire. The study was conducted in strict accordance with the ethical guidelines of the Declaration of Helsinki and was approved by the Human Research Ethics Committee at the Città della Salute e della Scienza Hospital (registration number 0077310). All participants were informed about the study and consented to participate. They were assured that participation was voluntary. The participants were also informed that their refusal to participate would not affect their care.

### Measures

The QLQ-C30 is composed of 30 items that measure six functional dimensions (emotional, physical, global health, cognitive, role and social) and eight symptoms (appetite loss, constipation, fatigue, nausea/vomiting, pain, diarrhea, dyspnea and insomnia); only one item is related to financial problems. The dimension scores range from 0 to 100. For the functional dimensions, a higher score represents a better quality of life, while for the symptoms scales, lower values indicate a better quality of life [27].

### Model specification and data analyses

#### *RS detection and the assessment of true change with the formative indicator model*

According to Boehmer and Luszczynska, physical symptoms were considered formative indicators [16]. More specifically, among the physical symptoms included in the EORTC QOL-C30 questionnaire, only fatigue (FA) and pain (PA) were used; they are also used as reflective indicators in the HRQoL questionnaire [28, 29], thus allowing us to compare the two alternative specifications. Three reflective indicators were considered to resolve the issue of model identification: the two items of the global health subscale, overall health (Q29, OH) and quality of life (Q30, QoL), and a global measure, which was the average of the remaining five functional subscale scores (FC). Hence, two formative and three reflective indicators were specified in the formative indicator model (Fig. 2).

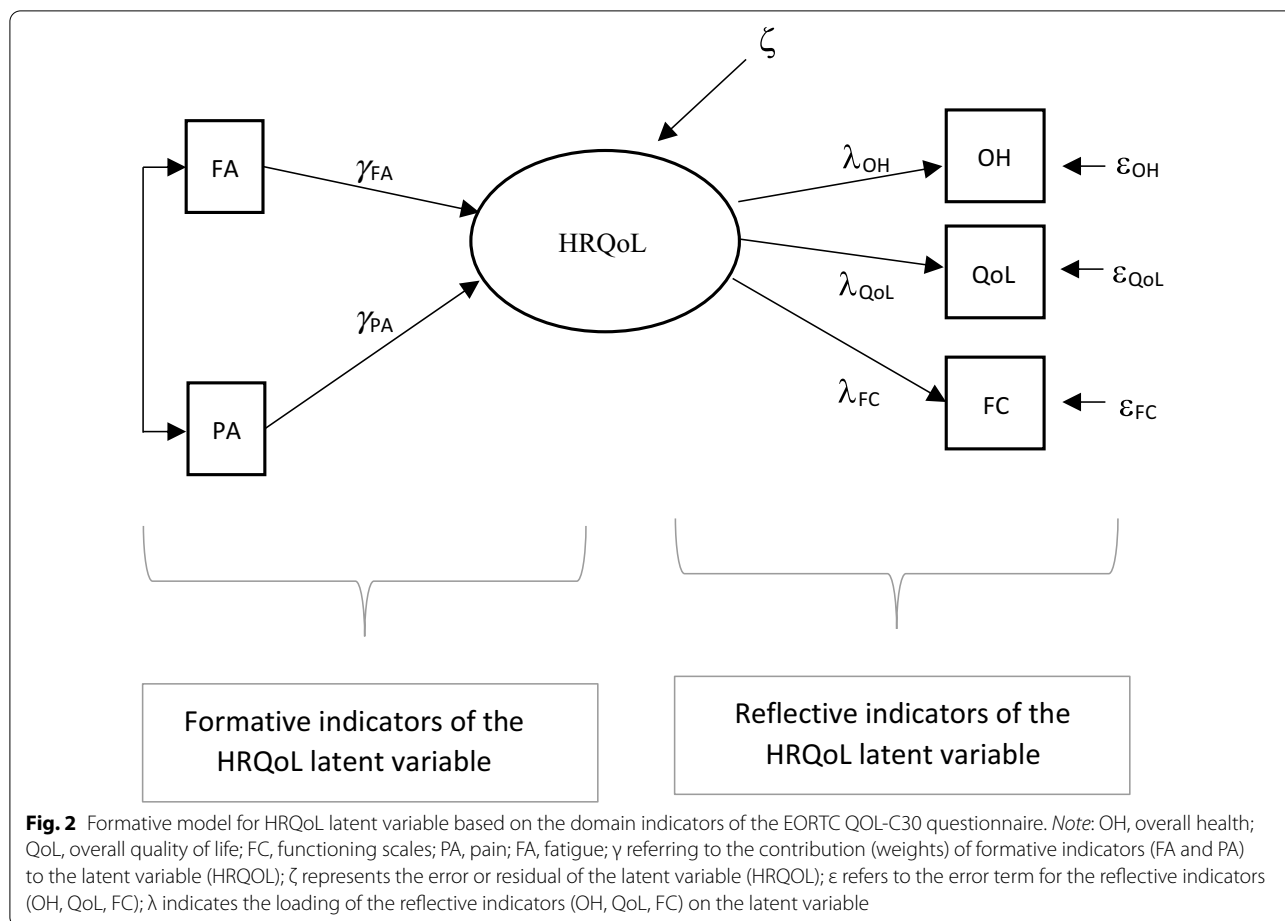
Before going into the details of the procedure, it is worth noting that the model specification depicted in Fig. 2 is the same as that of a multiple indicator multiple cause (MIMIC) model with the mean structure part included, although the parameters are interpreted differently [11]. A latent variable exists that receives some arrows and emits some others; however, while the emitted arrows have the same interpretation as reflective indicators that is they are observed expressions of the latent variable (multiple indicators), the receiving arrows are interpreted as formative indicators in a formative model (contributing to define the latent variable meaning) and as explanatory variables (multiple causes) in a MIMIC model. Formative indicators are conceptually different from explanatory variables because they are viewed as defining characteristics of the construct and eliminating one of them may alter the conceptual domain of the construct.

RS detection and the assessment of true change were performed in four steps.

*Step 1* The measurement invariance of the reflective indicators of HRQoL (OH, QoL and FC) that were included to resolve the identification issue was assessed. A baseline model with no equality constraints among the parameters between the two time points and a model with equality constraints on loadings, intercepts and error variances were estimated and compared. The fit of the baseline model was evaluated according to the following criteria: root mean square error of approximation (RMSEA)  $\leq 0.08$  and comparative fit index (CFI)  $\geq 0.95$  [30, 31]. The chi-square difference test ( $\Delta\chi^2$ ) was used to compare the fit of the two models.

*Step 2* The measurement invariance of the reflective indicators of the fatigue and pain constructs was assessed, which was averaged to create the two formative indicators (PA and FA). As in Step 1, the baseline model was compared to the constrained model. The same statistical analysis as that used in Step 1 was used to assess and compare the fit of the models.

*Step 3* For the assessment of RS, a baseline model in which the reflective part of the model was specified, as a result of Step 1, was established, and parameters of the formative part of the model were freely estimated (except for the mean of the latent variable at  $t=1$ , which was set to zero). Both intercepts were set to zero, whereas here and in the following models, the covariances between formative indicators were freely estimated. The baseline model was compared to a constrained model, in which the regression coefficients of the formative indicators and the variance of the latent variable were set to be equal between the two time points. The same statistical analysis as that used in Step 1 was used to assess and compare the fit of the models.



A statistically significant value for  $\Delta\chi^2$  was considered to be indicative of RS, and modification indices (MI) were used to identify the source of the lack of invariance.

*Step 4* Estimation of the final model, with relaxed untenable constraints, and an evaluation of the true change.

The four steps are summarized in Fig. 3.

**RS detection and assessment of true change with the reflective indicator model (Oort’s procedure)**

The five QLQ-C30 indicators previously described (FA, PA, OH, QoL and FC) were defined as reflective indicators of the latent HRQoL construct. For RS detection and the evaluation of true change, the original Oort’s procedure was used [6].

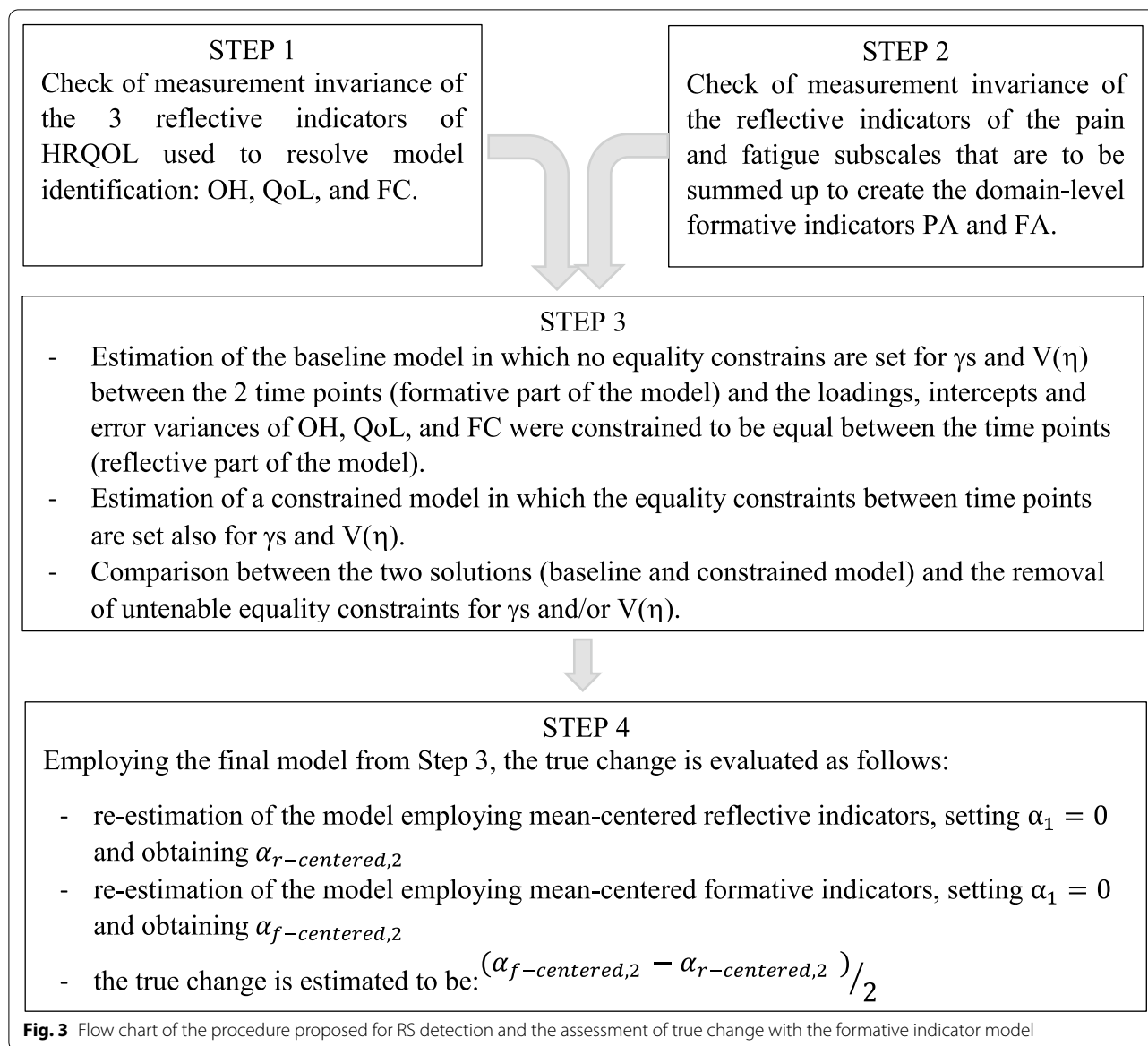
SEM analyses were performed under MPLUS v7.3. The robust maximum likelihood (MLR) method was used for the estimation because the normality assumption was violated.

**Results**

A cohort of 258 CRC patients completed the QLQ-C30 after diagnosis and 6 months later. The missing data rate was not greater than 2% for any of the items. For the items included in the analysis, there were no missing data.

Two-thirds of the participants were diagnosed with colon cancer (N=162, 62.8%), the remaining participants had rectal cancer, approximately 80% of the participants underwent surgery (N=213, 82.6%), and 51.4% of the participants were treated with chemotherapy. Almost half of the patients were men (N=55.4%), and only 34.5% had a high school degree. The participants’ ages ranged from 30 to 89, with a mean of 67.5 (standard deviation 10.5).

Table 2 shows the baseline and follow-up descriptive statistics of the five QLQ-C30 indicators included in the SEM models.



**Fig. 3** Flow chart of the procedure proposed for RS detection and the assessment of true change with the formative indicator model

**RS detection and the assessment of the true change with the formative indicator model**

*Step 1* As shown at the top of Table 3, the baseline model for the reflective indicators of HRQoL (OH, QoL and FC) fit the data perfectly since every reflective model with three indicators is exactly identified. With regard to the constrained model, the fit was good (RMSEA = 0.04, CFI = 0.99), and the magnitude of worsening with respect to the baseline model was not statistically significant ( $\Delta\chi^2(8) = 13.13, p = 0.108$ ).

*Step 2* The baseline model for the formative indicators of HRQoL, i.e., those relating to the five items measuring pain and fatigue, fit the data well (RMSEA = 0.03,

CFI = 0.99). The magnitude of worsening in fit of the constrained model (with equality constraints on loadings, intercepts and uniqueness between the two time points) was statistically significant ( $\Delta\chi^2(11) = 44.56, p < 0.001$ ) but was not statistically significant after the equality constraints on two error variances were removed ( $\Delta\chi^2(9) = 13.01, p = 0.162$ ).

*Step 3* The RS baseline model performed well (RMSEA = 0.05, CFI = 0.98). When the invariance constraints were imposed on the regression coefficients and on the construct residual variance, the fit of the model was not statistically worse than that of the baseline model ( $\Delta\chi^2(3) = 3.59, p = 0.310$ ), leading to the final formative model in Fig. 4a.

**Table 2 Descriptive value of the QLQ-C30 indicators included in SEM models at baseline (T1) and at the six-month follow-up (T2)**

QLQ-C30 indicators included in SEM models	M	SD	Skew	Kurt
	T <sub>1</sub> ; T <sub>2</sub>	T <sub>1</sub> ; T <sub>2</sub>	T <sub>1</sub> ; T <sub>2</sub>	T <sub>1</sub> ; T <sub>2</sub>
Overall health (Q29, OH)	66.7; 70.7	23.3; 22.2	-0.47; -0.66	-0.11; 0.13
Overall QoL (Q30, QoL)	67.9; 73.0	24.2; 22.3	-0.54; -0.58	-0.43; -0.41
Functioning scales <sup>a</sup> (FC)	81.9; 85.8	14.7; 14.9	-1.50; -1.58	2.59; 2.65
Fatigue (FA)	30.7; 25.8	22.8; 24.7	.74; 1.03	.37; .69
Pain (PA)	17.6; 11.7	22.4; 20.2	1.30; 2.15	1.02; 4.73

<sup>a</sup> Average score of the five functioning sub-scales (physical, role, cognitive, social, emotional)

T1 = baseline; T2 = six-month follow-up

*Step 4* As the invariance condition was fulfilled in Step 3, no RS was present according to the formative indicator model. To obtain evidence about true change in the HRQoL scores, the model intercepts were considered. In the model in which the formative indicators were mean centered, the intercept at t = 2 ( $\alpha_{f\_centered,2}$ ) was 4.00 (p < 0.001), and in the model in which the reflective indicators were mean centered, the intercept at t = 2 ( $\alpha_{r\_centered,2}$ ) was -3.05 (p < 0.001). The estimated true change was:

$$(\alpha_{f\_centered,2} - \alpha_{r\_centered,2})/2 = (4.00 + 3.05)/2 = 3.53.$$

**RS detection and assessment of true change with the reflective indicator model (Oort’s procedure)**

In the Oort’s procedure baseline model, we imposed the same invariance constraints relating to the three indicators OH, QoL and FC, as we did in Step 3 of our procedure, to guarantee the comparability of the results between the two RS procedures. As shown in Table 4, the baseline model fit well with the data (RMSEA = 0.04, CFI = 0.99), but when invariance constraints were imposed on the factor loadings, intercepts and residual variances relative to the FA and PA indicators, the magnitude of worsening in fit was quite significant ( $\Delta\chi^2(7) = 14.45, p = 0.044$ ).

Based on MI values and theoretical considerations, the equality constraint on the PA loading was relaxed (Fig. 4b), obtaining a model fit that was not significantly different from that of the baseline:  $\Delta\chi^2(6) = 9.45, p = 0.150$ . The loading of the pain indicator was -1.25 at t = 1 and -0.97 at t = 2, meaning that pain was a better indicator of HRQoL at t = 1. In terms of RS detection, this result suggests that a reprioritization has occurred: at t = 1 pain seems to be a more important component of HRQoL than during follow-up. The estimated true change with respect to the last model was 3.87.

**Discussion**

In this study, a procedure was proposed to detect RS when the indicators of the target construct are formative. According to this proposal, measurement models with formative indicators allow *reconceptualization* and

**Table 3 Fit statistics for the formative indicator models**

Model	$\chi^2$	df	p-value	RMSEA	CFI	$\Delta\chi^2$	$\Delta df$	p-value
Step 1: invariance of reflective indicators (OH, QoL, FC) used for identification Purpose								
Baseline	2.59	3	0.458	0	1	-	-	-
Constrained model <sup>a</sup>	16.06	11	0.139	0.04	0.99	13.13	8	0.108
Step 2: invariance of items used to create the two domain-level formative indicators (PA, FA)								
Baseline	30.30	24	0.175	0.03	0.99	-	-	-
Constrained model <sup>b</sup>	74.14	35	<0.001	0.07	0.97	44.56	11	<0.001
Constrained model <sup>c</sup>	43.26	33	0.109	0.03	0.99	13.01	9	0.162
Step 3: RS detection—invariance in the formative indicator model								
Baseline	51.59	30	0.01	0.05	0.98	-	-	-
Constrained model <sup>d</sup>	54.80	33	0.01	0.05	0.98	3.59	3	0.310

Step 1–step 3 are illustrated in Fig. 3

OH, overall health; QoL, overall quality of life; FC, functioning scales; PA, pain; FA, fatigue; RMSEA, root mean square approximation; CFI, comparative fit index

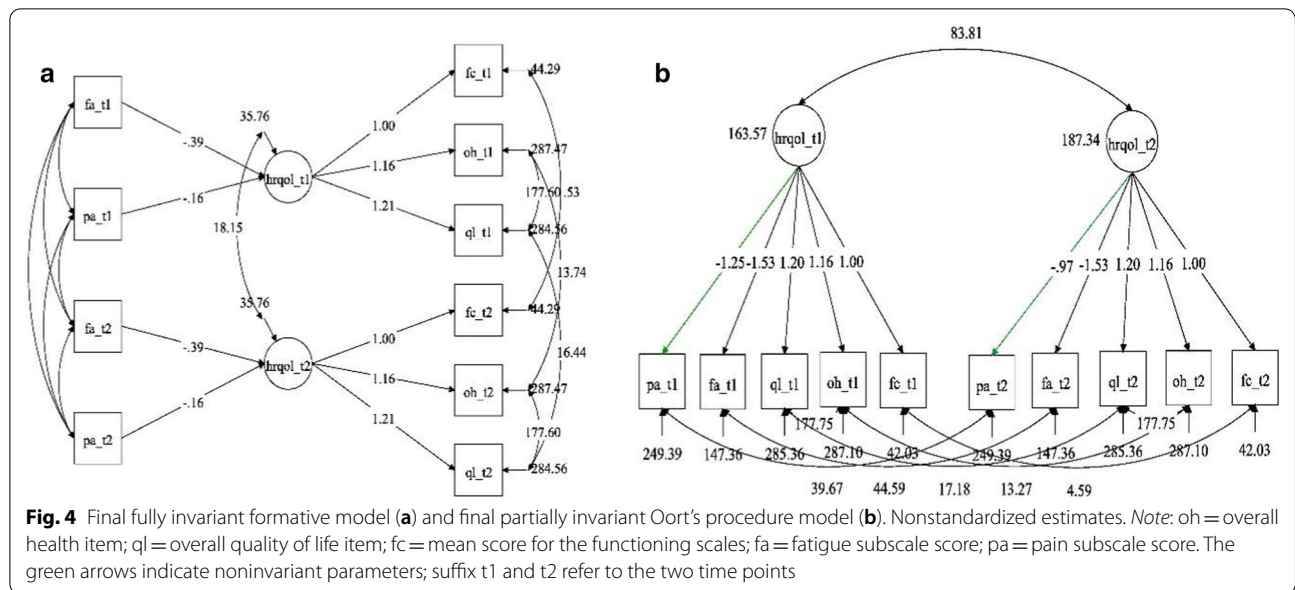
<sup>a</sup> Loadings, intercepts and residual variances of the three indicators, OH, QoL and FC, are constrained to be equal across the two time points; the mean of the latent variable is set to zero at t = 1 and freely estimated at t = 2

<sup>b</sup> Loadings, intercepts and residual variances of the five items measuring the latent variables PA and FA are constrained to be equal across the two time points; the means of the latent variables are set to zero at t = 1 and freely estimated at t = 2

<sup>c</sup> After removing two error variance equality constraints

<sup>d</sup> PA and FA regression coefficients and the unexplained variance of the latent variable is constrained to be equal across the two time points and the model intercept is set to zero at t = 1 and freely estimated at t = 2





**Table 4** Fit statistics for the reflective indicator models

Model	$\chi^2$	df	p-value	RMSEA	CFI	$\Delta\chi^2$	$\Delta$ df	p-value
Baseline	51.54	34	0.027	0.04	0.99	–	–	–
Constrained model <sup>a</sup>	66.48	41	0.007	0.05	0.98	14.45	7	0.044
Constrained model <sup>b</sup>	61.05	40	0.018	0.04	0.98	9.45	6	0.150

df, degree of freedom; RMSEA, root mean square approximation; CFI, comparative fit index

<sup>a</sup> Loadings, intercepts and residual variances of the five indicators, OH, QoL and FC, PA and FA, are constrained to be equal across the two time points; the mean of the latent variable is set to zero at t = 1 and freely estimated at t = 2

<sup>b</sup> After removing the equality constraints between the PA loadings

reprioritization forms of RS to be detected, whereas recalibration is not an issue; the intercepts and residual variance of the formative indicators are not model parameters. The true change in HRQoL can be estimated on the basis of two estimations of the model intercept, once by mean centering the variables used to solve the identification issue and once by mean centering the formative indicators. A measurement model with only formative indicators is not identified; to solve the problem, the target latent variable should emit at least two paths. These two paths can be related to two formative indicators or two other latent or observed uncorrelated variables that depend on the target latent variable. According to Jarvis et al. (2003), whenever possible, it is preferable to use two or more reflective indicators as “(a) the formative construct is identified on its own and can go anywhere in the model, (b) one can include it in a confirmatory factor model and evaluate its discriminant validity and measurement properties, and (c) the measurement parameters should be more stable and less sensitive to changes in the

structural relationships emanating from the formative construct” (p. 213).

Our procedure can be applied when the formative indicators are at the domain level and each domain indicator is measured by two or more time-invariant reflective indicators. This is because the mean scores of formative indicators contribute directly to the formation of the latent variable mean score and to the evaluation of true change across time points. Hence, they should not show uniform recalibration.

The procedural steps proposed for the assessment of reconceptualization and reprioritization are the same as those used by Diamantopoulos and Papadopoulos [16] for assessing the cross-nation invariance of formative indicators, whereas the last step regarding the mean structure part of the model is a completely novel proposal but essential when assessing true change in a construct of interest.

In the empirical example, the newly proposed procedure was applied to a model with three reflective indicators and two formative indicators of the latent variable,

HRQoL, and this procedure was compared to the traditional procedure applied to the model, in which all five indicators of HRQoL were modeled as reflective indicators. As a result, the estimated true change values for the proposed and traditional models were quite similar, as they were 3.53 and 3.87, respectively, but in terms of a lack of measurement invariance across two measurement occasions, the results differed. Defining fatigue and pain as formative indicators led to parameter estimates that were time invariant, whereas when the two indicators were defined as reflective, the loading of pain was not invariant between the two time points, and this lack of invariance could be interpreted as RS. In terms of these findings, it suggests the presence of reprioritization: pain was less important at the follow-up than at baseline. Thus, defining the indicators as formative or reflective can lead to different results in terms of the measurement invariance across time points and thus in terms of RS assessment.

It is critical that the appropriate model specification (reflective vs formative) is selected because measurement mis-specification leads to biased parameter estimates [15]. Whether an indicator is reflective or formative can be evaluated by means of empirical tests [32] or established on the basis of theoretical grounds, as the nature of the indicators depends on the definition of HRQoL that the researcher is adopting [33].

Regarding the limitations of the study, the new procedure was tested in only a single cohort and a single HRQoL questionnaire; it is not possible to generalize our findings of a few differences between the two RS detection approaches, as the results may be specific to the cohort and measures included in this study. The inclusion of additional external criteria for assessing RS would have helped us ascertain which of the two models was more specific since the fit indices of the two baseline models were quite similar. Moreover, future simulation studies are required to determine the validity of the proposed method.

In this study, RS was assessed from a measurement perspective, and future work is needed to integrate the presence of formative indicators within SEM models that investigate RS from a conceptual perspective [34, 35]. The HRQoL latent construct and its indicators (reflective and formative) can be analyzed in a network of relationships with other constructs or sociodemographic variables to detect possible bias both in the explanation and in the measurement of HRQoL.

## Conclusion

An assumption in the use of HRQoL questionnaires is that they are invariant over time, i.e., they are able to capture the real change in the interviewees even in

presence of the psychological adjustments that can occur when people are facing novel situations. This means that responses to HRQoL measures over time may vary not only because health or quality of life has changed but also because people may have changed their perception of what health or quality of life means to them (RS). In assessing RS and true change by means of the SEM approach, some indicators are better conceived as reflective, whereas others as better conceived as formative. The current literature does not provide methodological tools for assessing true changes in HRQoL when the measurement model includes formative indicators. The procedure proposed in this study aims to fill this gap.

## Acknowledgements

The authors are grateful to Laura Fiorini and Laura Davico for their valuable help in administering the questionnaires to patients.

## Authors' contributions

The study is based on an idea developed by RR who participated in the statistical analyses (execution, organization, review and critique) and participated in writing the paper. ST shared the project idea, performed the statistical analyses (execution, organization and discussion), and participated in writing the paper. DC shared the project idea, performed the statistical analyses (execution, organization and discussion), and participated in writing the paper. PR, GR, LF and SB assisted with the study design, data collection and writing the manuscript. All the authors discussed the results and read and approved the final manuscript.

## Funding

The study has not received funding.

## Availability of supporting data

Not applicable.

## Ethics approval and consent to participate

The study was conducted in strict accordance with the ethical guidelines in the Declaration of Helsinki and was approved by the Human Research Ethics Committee at "Città della Salute e della Scienza" hospital (registration number 0077310). All participants were informed about the study and consented to participation.

## Consent for publication

Not applicable.

## Competing interests

All the co-authors declare no financial competing interests nor any relevant conflicts of interests associated with this work.

## Author details

<sup>1</sup> Department of Human and Social Sciences, University of Aosta Valley, Aosta, Italy. <sup>2</sup> Department of Psychology, University of Turin, Turin, Italy. <sup>3</sup> Unit of Cancer Epidemiology, "Città Della Salute E Della Scienza" Hospital, University of Turin, CPO Piemonte, Turin, Italy. <sup>4</sup> SSD Colorectal Cancer Unit, Dipartimento Di Oncologia, "Città Della Salute E Della Scienza Di Torino" Hospital, Turin, Italy.

Received: 23 July 2020 Accepted: 25 December 2020

Published online: 06 January 2021

## References

1. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol*. 2003;56(5):395–407.

2. Sajobi TT, et al. Scoping review of response shift methods: current reporting practices and recommendations. *Qual Life Res.* 2018;27(5):1133–46.
3. Vanier A, et al. Overall performance of Oort's procedure for response shift detection at item level: a pilot simulation study. *Qual Life Res.* 2015;24(8):1799–807.
4. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med.* 1999;48(11):1507–15.
5. 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.
6. Oort FJ. Using structural equation modeling to detect response shifts and true change. *Qual Life Res.* 2005;14(3):587–98.
7. Rapkin BD, Schwartz CE. Advancing quality-of-life research by deepening our understanding of response shift: a unifying theory of appraisal. *Qual Life Res.* 2019;28(10):2623–30.
8. Schwartz CE, Sprangers MA. Reflections on genes and sustainable change: toward a trait and state conceptualization of response shift. *J Clin Epidemiol.* 2009;62(11):1118–23.
9. Fleuren BP, et al. Handling the reflective-formative measurement conundrum: a practical illustration based on sustainable employability. *J Clin Epidemiol.* 2018;103:71–81.
10. Bollen KA, Diamantopoulos A. In defense of causal-formative indicators: a minority report. *Psychol Methods.* 2017;22(3):581.
11. Jarvis CB, MacKenzie SB, Podsakoff PM. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *J Consum Res.* 2003;30(2):199–218.
12. Oort FJ. Towards a formal definition of response shift (in reply to GW Donaldson). *Qual Life Res.* 2005;14:2353–5.
13. Donaldson GW. Structural equation models for quality of life response shifts: promises and pitfalls. *Qual Life Res.* 2005;14(10):2345–51.
14. Bollen KA, Davis WR. Causal indicator models: Identification, estimation, and testing. *Struct Equ Model Multidiscip J.* 2009;16(3):498–522.
15. Bollen K. *Structural equations with latent variables.* New York: Wiley; 1989.
16. Boehmer S, Luszczynska A. Two kinds of items in quality of life instruments: "indicator and causal variables" in the EORTC qlq-c30. *Qual Life Res.* 2006;15(1):131–41.
17. Bollen K, Lennox R. Conventional wisdom on measurement: a structural equation perspective. *Psychol Bull.* 1991;110(2):305–14.
18. Fayers PM, Hand DJ. Factor analysis, causal indicators and quality of life. *Qual Life Res.* 1997;6(2):139–50.
19. Traa MJ, et al. Evaluating quality of life and response shift from a couple-based perspective: a study among patients with colorectal cancer and their partners. *Qual Life Res.* 2015;24(6):1431–41.
20. Bernhard J, et al. Quality of life as subjective experience: Reframing of perception in patients with colon cancer undergoing radical resection with or without adjuvant chemotherapy. *Ann Oncol.* 1999;10(7):775–82.
21. Ito N, et al. Response shift in quality-of-life assessment in patients undergoing curative surgery with permanent colostomy: a preliminary study. *Gastroenterol Nurs.* 2010;33(6):408–12.
22. Neuman HB, et al. Rectal cancer patients' quality of life with a temporary stoma: shifting perspectives. *Dis Colon Rectum.* 2012;55(11):1117–24.
23. Fayers PM, et al. Causal indicators in quality of life research. *Qual Life Res.* 1997;6(5):393–406.
24. Giesinger JM, et al. Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. *J Clin Epidemiol.* 2016;69:79–88.
25. Bollen KA. Evaluating effect, composite, and causal indicators in structural equation models. *Mis Q.* 2011;32:359–72.
26. Bollen KA, Bauldry S. Three Cs in measurement models: causal indicators, composite indicators, and covariates. *Psychol Methods.* 2011;16(3):265–84.
27. Aaronson NK, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *JNCI J Natl Cancer Inst.* 1993;85(5):365–76.
28. Vickrey B, et al. A health-related quality of life measure for multiple sclerosis. *Qual Life Res.* 1995;4(3):187–206.
29. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care.* 1992;30:473–83.
30. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model Multidiscip J.* 1999;6(1):1–55.
31. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. *Testing structural equation models.* Newbury Park: Sage; 1993. p. 136–92.
32. Bollen KA, Ting K-F. A tetrad test for causal indicators. *Psychol Methods.* 2000;5(1):3–22.
33. Costa DS. Reflective, causal, and composite indicators of quality of life: a conceptual or an empirical distinction? *Qual Life Res.* 2015;24(9):2057–65.
34. Oort FJ, Visser MR, Sprangers MA. Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. *J Clin Epidemiol.* 2009;62(11):1126–37.
35. King-Kallimanis BL, et al. Structural equation modeling of health-related quality-of-life data illustrates the measurement and conceptual perspectives on response shift. *J Clin Epidemiol.* 2009;62(11):1157–64.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

