Qual Life Res (2014) 23:2531–2543
DOI 10.1007/s11136-014-0693-9

Measurement of stable changes of self-management skills after rehabilitation: a latent state–trait analysis of the Health Education Impact Questionnaire (heiQ™)

M. Schuler · G. Musekamp · J. Bengel · M. Schwarze · K. Spanier · Chr. Gutenbrunner · I. Ehlebracht-König · S. Nolte · R. H. Osborne · H. Faller

Accepted: 9 April 2014 / Published online: 29 April 2014

Abstract

Purpose To assess stable effects of self-management programs, measurement instruments should primarily capture the attributes of interest, for example, the self-management skills of the measured persons. However, measurements of psychological constructs are always influenced by both aspects of the situation (states) and aspects of the person (traits). This study tests whether the Health Education Impact Questionnaire (heiQ™), an instrument assessing a wide range of proximal outcomes of self-management programs, is primarily influenced by person factors instead of situational factors. Furthermore, measurement invariance over time, changes in traits and predictors of change for each heiQ™ scale were examined.

Methods Subjects were $N = 580$ patients with rheumatism, asthma, orthopedic conditions or inflammatory bowel disease, who filled out the heiQ™ at the beginning, the end of and 3 months after a disease-specific inpatient rehabilitation program in Germany. Structural equation modeling techniques were used to estimate latent trait-change models and test for measurement invariance in each heiQ™ scale. Coefficients of consistency, occasion specificity and reliability were computed.

Results All scales showed scalar invariance over time. Reliability coefficients were high (0.80–0.94), and consistency coefficients (0.49–0.79) were always substantially higher than occasion specificity coefficients (0.14–0.38), indicating that the heiQ™ scales primarily capture person factors. Trait-changes with small to medium effect sizes were shown in five scales and were affected by sex, age and diagnostic group.

Conclusion The heiQ™ can be used to assess stable effects in important outcomes of self-management programs over time, e.g., changes in self-management skills or emotional well-being.

Keywords Self-management · Assessment · Latent state–trait theory · Latent trait-change model · Measurement invariance · Chronic disease

Electronic supplementary material The online version of this article (doi:10.1007/s11136-014-0693-9) contains supplementary material, which is available to authorized users.

M. Schuler (✉) · G. Musekamp · H. Faller
Department of Medical Psychology, Medical Sociology, and Rehabilitation Sciences, University of Würzburg, Klinikstr. 3, 97070 Würzburg, Germany
e-mail: m.schuler@uni-wuerzburg.de

J. Bengel
Department of Psychology, University of Freiburg, Freiburg, Germany

M. Schwarze · K. Spanier · Chr. Gutenbrunner
Hospital for Rehabilitation Medicine, Medical School Hannover, Hannover, Germany

I. Ehlebracht-König
Rehabilitation Center Bad Eilsen, Bad Eilsen, Germany

S. Nolte
Department of Psychosomatic Medicine, Medical Clinic, Charité – Universitätsmedizin Berlin, Berlin, Germany

S. Nolte · R. H. Osborne
School of Health and Social Development, Population Health Strategic Research Centre, Deakin University, Melbourne, Australia

Introduction

Self-management programs for chronic diseases seek to generate substantial positive change in individuals in areas such as self-management abilities, health-directed behavior or illness-related emotional distress [1–4]. "Stable changes" means that observed variations over time in these attributes do not primarily depend on specific (measurement) situations, but on "real" changes in the persons. However, measurements of psychological attributes are always influenced by aspects of both the measurement situation and aspects of the measured person, albeit to varying degrees [5]. Instruments to measure situation-independent constructs should therefore primarily capture person factors and should be little affected by situational factors. Otherwise, it would be difficult to differentiate between mere situation-dependent and short-term *variability* and situation-independent long-term *changes* in the constructs of interest [6, 7].

A framework to empirically test assumptions about the influences of either measurement situations or measured persons is offered by latent state–trait theory (LST theory) [5, 8, 9]. LST theory postulates that the variance of an observed variable (OV) can be broken down into a trait component, a state component and measurement error. The trait component reflects the influence of the measured person, whereas the state component reflects the influence of measurement situations (and person × situation interactions) on the OV. If method effects of different measures for the same construct are taken into account, method factors also explain parts of the variance of the OV. From these theoretical assumptions, psychometric parameters, referred as LST parameters, can be defined that quantify influences of these different sources [5, 8, 10]. For example, in the SF-36, it was shown that 69–75 % of the variance in the mental component summary, which measures situation-independent mental health, is explained by stable person factors [11], while up to 77 % of the variances in measures of mood states can be attributed to measurement situations [12].

A variety of different LST models can be tested empirically via structural equation modeling techniques [9]. Basically, models that assume a constant trait over time can be distinguished from models that allow trait-changes over time [13, 14]. In this study, a latent trait-change model is used to model changes in important outcomes of self-management programs in the context of inpatient rehabilitation.

A generic instrument that comprehensively captures goals of self-management programs is the Health Education Impact Questionnaire (heiQ™) [15, 16]. The heiQ™ measures proximal outcomes of self-management programs across eight different constructs, ranging from self-management skills to navigation in the health care system. The original English version as well as translations into German and French show good psychometric properties in terms of factorial validity, concurrent validity and reliability [15, 17, 18]. However, to be a valid instrument to measure stable changes as a result of attending a self-management program, the heiQ™ must also show to be primarily influenced by the measured persons, not by the measured situations.

Therefore, this study seeks to answer the following main question: Are the items and scales of the heiQ™ primarily influenced by the measured persons (traits) or by the measurement situations/interactions of person × situations (states)?

Furthermore, the study explored how the traits measured by the heiQ™ scales change after inpatient rehabilitation that included self-management programs and whether these changes are associated with disease group, sex, age or initial trait values at the beginning of the inpatient rehabilitation.

Methods

Sample

All analyses are based on a subsample ($N = 580$) of subjects from of a study that translated and psychometrically tested the heiQ™ in Germany [18]. Patients with rheumatic disorders (e.g., psoriatic arthritis, ankylosing spondylitis; $n = 186$), asthma ($n = 170$), orthopedic conditions (e.g., chronic back pain; $n = 121$) and inflammatory bowel disease ($n = 103$) from five German rehabilitation clinics were included. All participants filled out the heiQ™ at the beginning (T1), the end of (T2) and 3 months after (T3) a disease-specific inpatient rehabilitation. All interventions lasted 3–4 weeks and included a self-management program aimed at enhancing self-management skills and health-directed behaviors as well as reducing emotional distress. Further details about the interventions are presented elsewhere [19–23] and are available upon request.

Measures

heiQ™

The heiQ™ contains 40 items (4-point response scale) across eight independent scales: *Positive and active engagement in life*, *Health-directed activities*, *Skill and technique acquisition*, *Constructive attitudes and approaches*, *Self-monitoring and insight*, *Health service navigation*, *Social integration and support*, and *Emotional distress*. The scale scores are formed by computing the mean of respective items. Generally, higher values in the heiQ™ scales indicate better status, except for

Emotional distress, in which higher values indicate higher distress [15, 16, 18].

Statistical analysis

The statistical analyses were done in three steps. First, for each heiQ™ scale, we tested whether factor structure and model parameters (e.g., factor loadings) did not change over time, i.e., whether the scales showed measurement invariance over time [24, 25]. At minimum scalar invariance (i.e., same factor structure and constant factor loadings/intercepts over time) is a prerequisite to interpret LST models accurately (Geiser et al., under review). Second, the change process was modeled for each heiQ™ scale using LST models as described below. Third, psychometric LST parameters were computed.

Measurement invariance over time

Configural, metric and scalar invariance over time [24, 26, 27] were tested through several confirmatory factor analyses. First, a measurement model over time—also called multi-state model in LST theory [9]—was computed for each heiQ™ scale (Fig. 1). All observed items OV_{it} of a measurement occasion t load on a latent state factor (S_t). S_1–S_3 were allowed to correlate with each other. They represent the common variance of the items of a scale on occasion t, and the correlations between the S_t account for stable individual differences over time. However, state- and trait-aspects are not clearly separated in a multi-state model [9]. Item-specific associations over time were modeled according to the correlated trait-correlated method minus one (CT − C(M − 1)) approach [28, 29]. In this approach, latent method factors were modeled for each indicator, except for a reference indicator (see below). The method factors represent stable aspects of each indicator that are not shared by the reference indicator.

To identify the model and test for configural invariance, variances of S_1–S_3 were set to 1 and corresponding mean values to 0 [30]; all other parameters were estimated freely. Metric invariance was tested by holding the factor loadings on the latent state variables (δ_{it}) and the factor loadings on the method factors (γ_{it}) constant over time, i.e., $\delta_{it} = \delta_i$ and $\gamma_{it} = \gamma_i$. The variance of S_1 was still fixed to 1, but the variances of S_2 and S_3 were now estimated freely. Scalar invariance was tested by additionally holding all intercepts constant over time. The mean value of S_1 was still fixed to zero, while the mean values of S_2 and S_3 were estimated freely.

Non-invariant parameters were identified via expected parameter changes (EPC) and modification indices using the software JruleMplus [31]. JruleMplus tests whether an

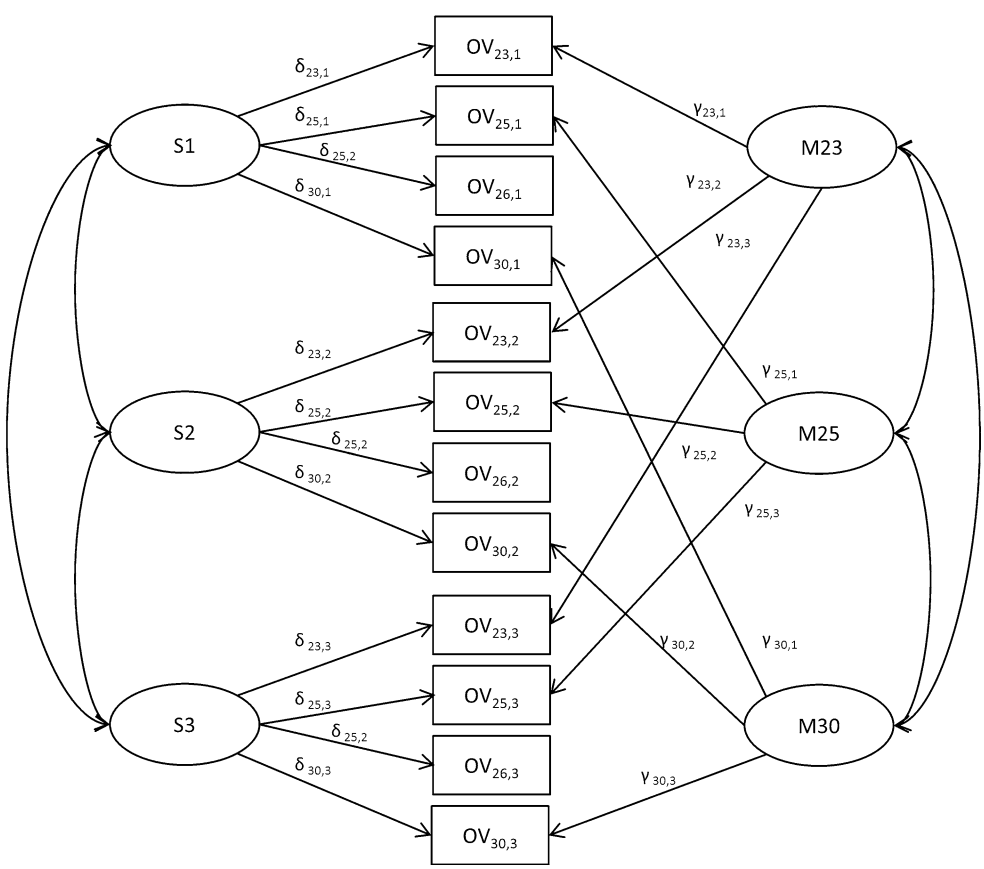

Fig. 1 Multi-state model of heiQ™ scale *Skill and technique acquisition* with three measurement occasions ($OV_{i,t}$: observed item i on time t; S_t: latent state variable of measurement occasion t; M_i: method factor of item i; δ_{it}: factor loading of item i on occasion t on S_t; γ_{it}: factor loading of item i on occasion t on method factor M_i)

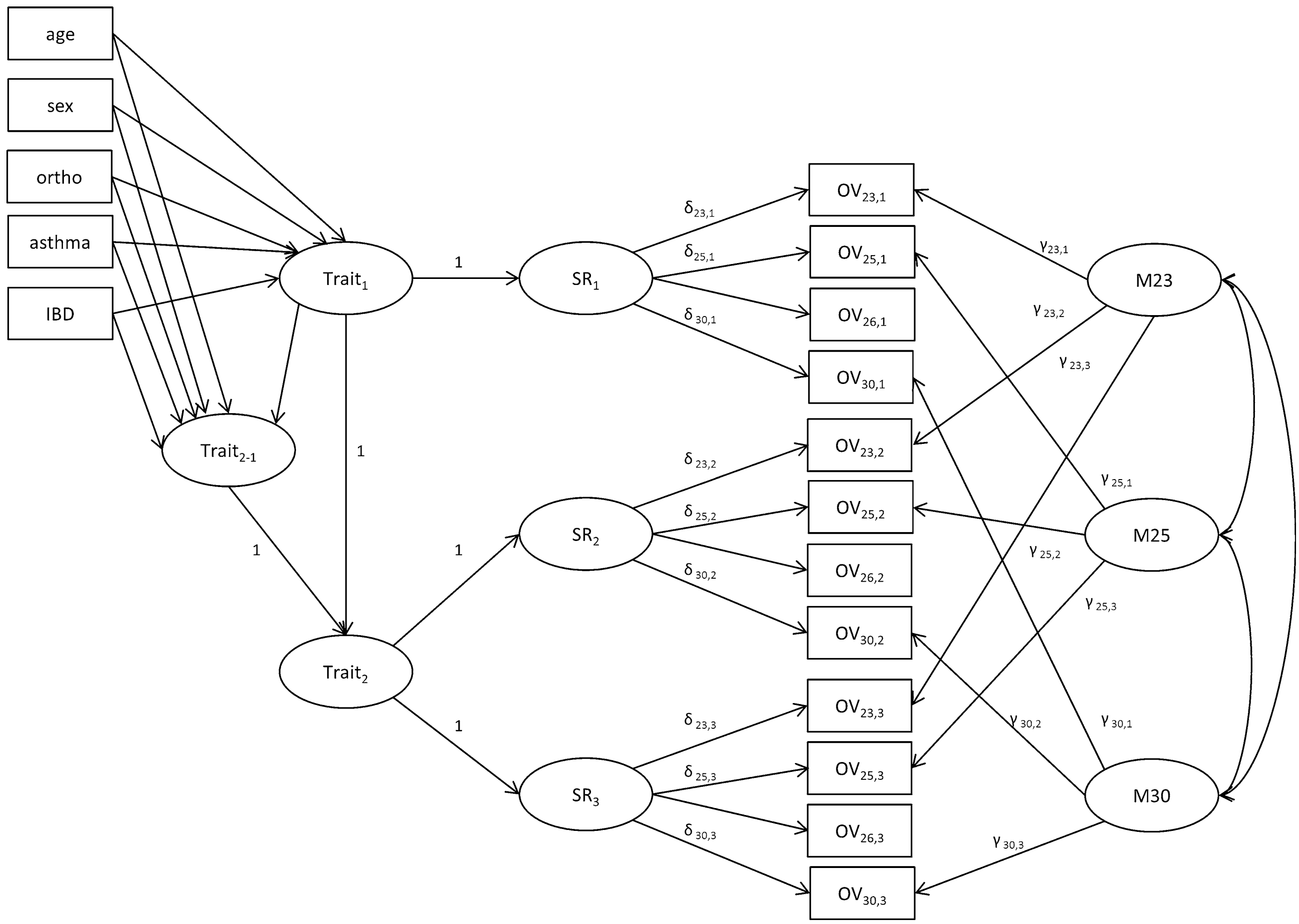

Fig. 2 Latent trait-change model of heiQ[TM] scale *Skill and technique acquisition* with three measurement occasions and predictors ($OV_{i,t}$: observed item *i* on time *t*; SR_t: latent state residual variable of measurement occasion *t*; M_i: method factor of item *i*; $Trait_k$: latent trait variable *k*; $Trait_{2\text{-}1}$: latent difference variable; δ_{it}: factor loading of item *i* on occasion *t* on S_t; γ_{it}: factor loading of item *i* on occasion *t* on method factor M_i; *ortho* orthopedic; *IBD* inflammatory bowl disease)

EPC exceeds a reference value. However, up to now, there are no guidelines for choosing appropriate reference values in invariance testing. In some studies, differences in intercepts about 0.3 lead to biased estimates in mean values, while differences in factor loadings about 0.2 did not [32; Schuler et al., in press]. Therefore, reference values were fixed to 0.25 in this study, representing the minimal difference in factor loadings or intercepts over time regarded as meaningful [33, 34]. Whenever a non-invariant parameter was identified, this parameter was set free and partial invariance models [35] were tested.

Estimation of LST model

Afterwards, each heiQ[TM] scale was modeled by a LST model (Fig. 2). The latent state variables S_t of Fig. 1 were separated in latent trait variables ($Trait_1$, $Trait_2$) and latent state residual variables (SR_1–SR_3). $Trait_1$ and $Trait_2$ represent the influence of the persons on the OV_{it} that is independent of the measurement situation, while SR_1–SR_3 represent the "pure" influence of the three measurement situations. Therefore, SR_1–SR_3 do not correlate anymore with each other.

To identify SR_1–SR_3, the factor loading of a reference indicator was fixed to 1 and the intercept was fixed to 0 on each measurement occasion. Indicators that best represented the content of the scale and that showed high factor loading as shown in [18] were chosen as reference indicator. All other factor loadings and intercepts were freely estimated, but should show scalar invariance over time.

SR_1 loaded only on $Trait_1$, while SR_2 and SR_3 loaded only on $Trait_2$. The rationale for this model is as follows: It was assumed that participating in a self-management program and, for example, learning new techniques to cope with the chronic illness may change the traits of the heiQ[TM] constructs between T1 and T2. This means that possible mean changes in heiQ[TM] items/scales between T1 and T2/T3 may indicate "real" trait-changes. But changes in heiQ[TM] items/scales between T2 and T3 were regarded as mere fluctuations around a common trait. For example, a decline in mean values between T3 and T2 may rather reflect an adaption process of new learned techniques on daily life circumstances than a "real" change in the measured trait. Note that these assumptions are difficult to test empirically; they rather represent underlying interpretations made by the authors.

Of note, Trait_1 and Trait_2 were not fully identified with only three latent state residual variables. As an additional restriction, we therefore assumed constant variances of the latent state residual variances over time [14]. Furthermore, the intercepts of SR_1 and SR_3 were fixed to zero, but the intercept of SR_2 was estimated freely.

Changes in latent trait variables

To estimate the difference between the two trait variables directly, a latent difference variable $\text{Trait}_{2\text{-}1}$ was introduced [14, 36]. A latent difference variable represents the true difference of two latent variables, i.e., a difference without measurement error. Technically, it is defined and identified by fixing the paths from Trait_1 and $\text{Trait}_{2\text{-}1}$ to Trait_2 at 1 and the variance of Trait_2 at 0 [36]. A standardized estimate was computed by dividing the mean of $\text{Trait}_{2\text{-}1}$ by the standard deviation of Trait_1 [37, 38]. Furthermore, to compare these values with more common parameters, the difference between T3 and T1 in a manifest heiQ™ scale was computed and divided by the standard deviation of the scale value on T1. Standardized differences of 0.2/0.5/0.8 were regarded as small/medium/large in latent and observed scores [39].

Method effects were again modeled according to the CT − C(M − 1) approach [28, 29]. The latent method factors have expected values of zero and do not correlate with latent trait factors or the latent state residuals.

Predictors of change in latent trait variables

Age, sex and diagnostic group were included as predictors of Trait_1 and $\text{Trait}_{2\text{-}1}$ in the model using weighted effect-coding [40].

LST parameters for items

From the estimated parameters of the LST models, the following five psychometric parameters were computed (abbreviations used in the equations are explained in the following paragraph) [28]:

$$\text{CCO}(\text{OV}_{it}) = \frac{\lambda_{it}^2 \text{Var}(\text{Trait}_k)}{\text{Var}(\text{OV}_{it})} \quad (1)$$

$$\text{UCO}(\text{OV}_{it}) = I(i \neq r)\frac{\gamma_{it}^2 \text{Var}(\text{MF}_i)}{\text{Var}(\text{OV}_{it})} \quad (2)$$

$$\text{TCO}(\text{OV}_{it}) = \text{CCO}(\text{OV}_{it}) + \text{UCO}(\text{OV}_{it}) \quad (3)$$

$$\text{OSpe}(\text{OV}_{it}) = \frac{\delta_{it}^2 \text{Var}(\text{SR}_t)}{\text{Var}(\text{OV}_{it})} \quad (4)$$

$$\text{Rel}(\text{OV}_{it}) = \text{TCO}(\text{OV}_{it}) + \text{OSpe}(\text{OV}_{it}) \quad (5)$$

The common consistency coefficient (CCO) represents the degree of the variance of an OV that is stable over time and explained by the trait factor Trait_k ($k = 1, 2$). The uniform consistency coefficient (UCO) represents the part of the variance that is also stable over time, but not shared with the reference indicator. These two coefficients sum up to the total consistency coefficient (TCO), reflecting the total stable part of the variance of an OV. The part of the variance that is explained by the measurement situation and/or person × situation interaction is represented by the occasion specificity coefficient (OSpe). The sum of TCO and OSpe results in the reliability coefficient (Rel) that indicates the degree to which the observed variance is explained by systematic influences, i.e., not determined by measurement error. Reliability coefficients above 0.7/0.8/0.9 were regarded as acceptable/good/very good. The terms λ_{it} and δ_{it} represent the factor loading of an item i at time t on Trait_k and SR_t, respectively, while γ_{it} represent the factor loading of item i at time t on the latent method factor i (except for the reference indicators r that do not load on any method factor).

LST parameters for scales

LST parameters were also estimated for whole heiQ™ scales. Aggregated latent variables were computed for each scale on each measurement occasion [41, 42]. The aggregated equations developed by Eid and Diener [42] were used and adjusted (see Appendix). Note that all computations of LST parameters were based on model-implied estimates.

Software, missing values and alpha

All structural equation models were computed using Mplus v7.1 [43] with robust maximum likelihood estimator and were based on covariance matrices. Missing values (0–3 %) were handled using full-information maximum likelihood estimation. Models were evaluated using chi^2-test and fit-indices Comparative Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA), but as the chi^2-test is highly influenced by sample size, a significant test did not automatically lead to model rejection. Generally, model fit was regarded as acceptable with CFI close to 0.95 and RMSEA close to 0.06 [44]. However, invariance tests were guided by EPCs/modification indices and not by measures of fit for the whole model. Therefore, models with fit-indices below these values may sometimes also be regarded as acceptable. Alpha was fixed to 0.05 for all analyses.

Results

Sample

Fifty-eight percent of the patients were female; mean age was 48.3 years (SD = 9.1). Detailed sample characteristics are found in Table 1.

Measurement invariance over time

All heiQ™ scales showed measurement invariance over time in factor loadings of the latent state factors S_t and in all intercepts (Table 2). Therefore, they fulfilled the requirements for scalar invariance over time with respect to the latent state factors.

Estimation of LST models

Model fit of LST models were nearly the same as the scalar invariance models of Table 2 (not shown). However, in *Health-directed activities*, estimation of the LST models leads to a negative residual variance in $\text{Trait}_{2\text{-}1}$ [45]. After fixing this parameter on 0.01, the model could be estimated regularly.

LST parameter for scales

Estimated LST parameters for the scales are presented in Table 3. Reliability coefficients range from 0.81 to 0.94. In all scales, TCO values are clearly higher than OSpe values, indicating that the scales are influenced more strongly by stable person factors than by situational factors. Highest OSpe values are found in *Skill and technique acquisition* (0.23–0.27) and *Health-directed activities* (0.28–0.39), i.e., up to 40 % of the latter scale's variance can be attributed to the measurement situation. CCO varies between 0.37 and 0.75, showing that 75 % of the observed variance in *Emotional distress* can be attributed to the common trait, while in most other scales, the common trait accounts for about 50 % of the variances. Method-specific factors account for 4–5 % (*Emotional distress*) to up to about 15 % (*Positive and active engagement in life*, *Self-monitoring and insight*) of the scale variances.

LST parameters for items

Table 4 shows the range of LST parameters for each scale per measurement occasion (detailed results are available

Table 1 Sample characteristics

	Rheumatism (n = 186) n (%)	Asthma (n = 170) n (%)	Orthopedic condition (n = 121) n (%)	Inflammatory bowel disease (n = 103) n (%)	Total[a] (N = 580) n (%)
Sex					
Male	50 (26.9)	104 (61.2)	50 (41.3)	38 (36.9)	242 (41.7)
Female	136 (73.1)	66 (39.8)	71 (58.7)	65 (63.1)	338 (58.3)
Living with a partner					
Yes	132 (74.4)	136 (81.4)	94 (78.3)	65 (63.7)	427 (75.3)
No	46 (25.6)	31 (19.6)	26 (21.7)	37 (36.3)	140 (24.7)
Employment status					
Working	132 (62.2)	149 (88.3)	105 (87.5)	86 (83.3)	472 (82.1)
Unemployed	19 (10.4)	10 (5.9)	8 (6.7)	3 (2.9)	40 (6.9)
Pension	7 (3.8)	2 (1.2)	0 (0)	2 (1.9)	11 (1.9)
Other (housewife…)	25 (13.7)	8 (4.8)	7 (5.8)	12 (11.7)	52 (9.0)
Occupation					
Laborer	65 (35.7)	92 (54.8)	2 (1.7)	10 (9.7)	169 (29.4)
Clerk/civil servant	102 (56.0)	63 (37.5)	110 (91.7)	86 (83.5)	361 (62.0)
Self-employed	5 (2.7)	8 (4.8)	7 (5.8)	6 (5.8)	26 (4.5)
Other	10 (5.5)	5 (3.9)	1 (0.8)	1 (1.0)	17 (0.03)
Age [M (SD)]	49.2 (8.9)	48.0 (8.5)	51.7 (7.5)	43.0 (10.4)	48.3 (9.1)

[a] Due to missing data, not all subgroups add up to N = 580

Table 2 Measurement invariance analysis over time of all heiQ™ scales

Invariance type	Chi^2 (*df*)	*p*	CFI	RMSEA	EPC > δ
Positive and active engagement in life					
Configural	155.012 (70)	<0.001	0.969	0.046	
Metric	187.85 (86)	<0.001	0.962	0.045	
Partial metric	171.79 (85)	<0.001	0.968	0.042	M151_1
Scalar	315.42 (93)	<0.001	0.918	0.064	
Health-directed activities					
Configural	48.35 (39)	0.145	0.997	0.020	
Metric	77.69 (50)	0.007	0.990	0.031	M9_1
Partial metric	66.98 (49)	0.045	0.992	0.025	
Scalar	92.83 (55)	0.001	0.986	0.034	
Skill and technique acquisition					
Configural	62.22 (39)	0.005	0.989	0.034	
Metric	84.84 (51)	0.002	0.986	0.034	M30_3
Partial metric	74.39 (50)	0.014	0.990	0.029	
Scalar	122.05 (56)	<0.001	0.973	0.045	
Constructive attitudes and approaches					
Configural	79.45 (69)	0.183	0.997	0.016	
Metric	96.76 (85)	0.180	0.997	0.015	
Scalar	108.42 (93)	0.131	0.996	0.017	
Self-monitoring and insight					
Configural	121.77 (107)	0.156	0.995	0.015	
Metric	149.80 (126)	0.073	0.992	0.018	M11_1
Partial metric	145.59 (125)	0.101	0.993	0.017	
Scalar	196.82 (135)	<0.001	0.978	0.028	
Health service navigation					
Configural	81.84 (69)	0.138	0.996	0.018	
Metric	109.17 (85)	0.040	0.993	0.022	M29_3 M32_3 M33_1
Partial metric	84.51 (82)	0.403	0.999	0.007	
Scalar	111.06 (90)	0.065	0.994	0.020	
Social integration and support					
Configural	82.93 (69)	0.121	0.997	0.019	
Metric	115.97 (85)	0.014	0.994	0.025	M35_3
Partial metric	100.47 (84)	0.106	0.997	0.018	
Scalar	125.30 (92)	0.010	0.993	0.025	
Emotional distress					
Configural	157.02 (117)	0.009	0.994	0.024	
Metric	190.82 (127)	<0.001	0.990	0.023	M12_3
Partial metric	184.51 (126)	0.001	0.991	0.028	
Scalar	277.55 (136)	<0.001	0.977	0.042	

M*i_t*: method factor of item *i* at time *t*; *df*: degrees of freedom; CFI: Comparative Fit Index; RMSEA: Root Mean Square Error of Approximation; EPC > δ: items with expected parameter changes > reference value δ in a fixed parameter (e.g., fixed loading in metric invariance testing); δ was set to 0.25 for factor loadings and intercept (see text)

upon request). The range of the parameters can be viewed as an indicator of homogeneity. The smaller the range, the more homogenous are the items of that scale in respective scale parameter. For example, reliability coefficients show little variation in *Social integration and support*, indicating that these items hardly differ in view of reliability.

Table 3 Latent state–trait parameters for each heiQ™ scale

Occasion	CCO	UCO	TCO	OSpe	Rel
Positive and active engagement in life					
T1	0.52	0.14	0.65	0.17	0.82
T2	0.52	0.13	0.65	0.20	0.85
T3	0.52	0.13	0.65	0.20	0.84
Health-directed activities					
T1	0.49	0.01	0.58	0.28	0.86
T2	0.37	0.13	0.50	0.39	0.89
T3	0.36	0.13	0.49	0.38	0.88
Skill and technique acquisition					
T1	0.49	0.09	0.58	0.23	0.80
T2	0.50	0.10	0.60	0.27	0.87
T3	0.51	0.07	0.57	0.27	0.85
Constructive attitudes and approaches					
T1	0.64	0.05	0.69	0.20	0.89
T2	0.65	0.05	0.70	0.21	0.91
T3	0.65	0.05	0.70	0.21	0.91
Self-monitoring and insight					
T1	0.54	0.12	0.66	0.15	0.81
T2	0.52	0.14	0.66	0.20	0.85
T3	0.51	0.15	0.66	0.20	0.85
Health service navigation					
T1	0.59	0.07	0.66	0.21	0.87
T2	0.60	0.07	0.67	0.23	0.90
T3	0.62	0.04	0.66	0.24	0.89
Social integration and support					
T1	0.68	0.10	0.78	0.14	0.92
T2	0.64	0.11	0.75	0.17	0.93
T3	0.65	0.09	0.74	0.18	0.92
Emotional distress					
T1	0.72	0.05	0.77	0.14	0.91
T2	0.75	0.05	0.79	0.15	0.94
T3	0.75	0.04	0.79	0.15	0.94

CCO common consistency, *UCO* unique consistency, *TCO* total consistency, *OSpe* occasion specificity, *Rel* reliability

For nearly all items, TCO values are clearly higher than OSpe values. Most scales show considerable ranges in CCO. However, they differ in the amount of CCO of the items. For example, the range of CCO values is >.30 in *Emotional distress* as well as in *Self-monitoring and insight*. However, even the items with lowest CCO value in *Emotional distress* (CCO = 0.34) shares one-third of its variance with the reference indicator, while in *Self-monitoring and insight,* the item with the lowest CCO value (CCO = 0.12) shares only about 12 % with the reference indicator.

Reliability coefficients range between 0.38 (item 3, "As well as seeing my doctor, I regularly monitor changes in my health") and 0.85 (item 21, "If I think about my health, I get depressed"). There is no general decrease or increase in LST parameters over time. Overall, CCO, OSpe and Rel are lower for most items than for respective scales. However, UCO is sometimes higher on the item level than on the scale level.

Changes in latent trait variables

The standardized means of $Trait_{2\text{-}1}$ across all disease groups (Table 5) vary from 0.07 (*Social integration and support*) to 0.57 (*Skill and technique acquisition*, *Self-monitoring and insight*), indicating that in some constructs trait-changes of low to medium size can be detected. In five scales, the statistical significant intercept of SR_2 indicates mean changes between T2 and T3. The highest intercept value is found in *Health-directed activites*. Table 4 also shows standardized mean differences between T1 and T3 in observed heiQ™ scales ($ES_{Obs3\text{-}1}$). In *Positive and active engagement in life*, *Health-directed activities* and *Self-monitoring and insight* $ES_{Obs3\text{-}1}$ were substantially lower than the standardized difference between latent trait variables. In all other scales, both latent and observed differences were nearly equivalent.

Predictors of change in latent trait variables

Table 6 shows that in some scales $Trait_1$ and/or $Trait_{2\text{-}1}$ mean values depend on diagnosis group. For example, patients with rheumatic disorders show lower T1 mean values in most scales and also smaller increases in three scales (*Self-monitoring and insight*, *Positive and active engagement in life* and *Health-directed activities*). Asthma patients show higher means in $Trait_1$ and in $Trait_{2\text{-}1}$ in *Skill and technique acquisition* and *Self-monitoring and insight* as well as lower *Emotional distress* on T1 and higher decline to T2. Higher gain for asthma patients were also found in *Positive and active engagement in life* and in *Health-directed activities*. Furthermore, mean differences and $Trait_1$ means depend on age and sex, in some scales. For example, women as well as younger patients show higher $Trait_1$ values and a higher decrease over time in *Emotional distress*. Women also show higher gains in *Constructive attitudes and approaches* and in *Positive and active engagement in life.*

Discussion

By using models of LST theory, this study showed that the scales of the German version of the heiQ™ are invariant over time that they are reliable and that they are primarily influenced by person factors that are independent of the measurement situation. Therefore, the heiQ™ may be a

Table 4 Range of LST parameters of respective heiQ™ items in each heiQ™ scale

Occasion	CCO	UCO	TCO	OSpe	Rel
Positive and active engagement in life					
T1	0.17–0.39	0.20–0.31	0.39–0.58	0.06–0.12	0.49–0.67
T2	0.19–0.47	0.21–0.31	0.45–0.61	0.07–0.12	0.53–0.73
T3	0.21–0.41	0.21–0.30	0.41–0.59	0.08–0.12	0.53–0.68
Health–directed activities					
T1	0.23–0.42	0.07–0.23	0.41–0.49	0.13–0.24	0.54–0.74
T2	0.17–0.26	0.08–0.27	0.26–0.44	0.18–0.27	0.53–0.61
T3	0.16–0.25	0.08–0.25	0.25–0.41	0.13–0.27	0.51–0.58
Skill and technique acquisition					
T1	0.10–0.48	0.05–0.32	0.35–0.48	0.05–0.22	0.45–0.70
T2	0.10–0.51	0.05–0.48	0.40–0.60	0.07–0.28	0.43–0.81
T3	0.13–0.53	0.06–0.25	0.38–0.55	0.05–0.22	0.43–0.77
Constructive attitudes and approaches					
T1	0.33–0.50	0.06–0.22	0.46–0.56	0.10–0.16	0.57–0.69
T2	0.35–0.55	0.06–0.25	0.50–0.63	0.12–0.18	0.62–0.77
T3	0.32–0.49	0.06–0.22	0.45–0.56	0.11–0.16	0.56–0.67
Self-monitoring and insight					
T1	0.12–0.53	0.08–0.31	0.34–0.53	0.04–0.15	0.38–0.67
T2	0.14–0.49	0.08–0.32	0.37–0.53	0.05–0.19	0.48–0.68
T3	0.16–0.49	0.10–0.33	0.41–0.55	0.03–0.15	0.51–0.64
Health service navigation					
T1	0.19–0.53	0.07–0.32	0.47–0.53	0.07–0.19	0.58–0.71
T2	0.20–0.53	0.09–0.22	0.42–0.65	0.08–0.21	0.50–0.81
T3	0.21–0.57	0.08–0.22	0.42–0.57	0.07–0.19	0.49–0.75
Social integration and support					
T1	0.39–0.60	0.16–0.24	0.60–0.64	0.08–0.13	0.71–0.74
T2	0.36–0.62	0.15–0.31	0.61–0.66	0.10–0.17	0.71–0.79
T3	0.33–0.56	0.14–0.31	0.52–0.58	0.08–0.13	0.68–0.72
Emotional distress					
T1	0.34–0.65	0.07–0.17	0.46–0.65	0.07–0.13	0.58–0.77
T2	0.40–0.71	0.08–0.19	0.54–0.72	0.08–0.14	0.63–0.85
T3	0.40–0.70	0.08–0.20	0.53–0.70	0.08–0.14	0.60–0.84

CCO common consistency, *UCO* unique consistency, *TCO* total consistency, *OSpe* occasion specificity, *Rel* reliability

useful instrument to measure stable effects of self-management programs over time. This study further showed that short-time changes in self-management skills and emotional well-being after inpatient rehabilitation remain stable over a period of at least 3 months.

The LST models of seven scales showed good fit values, suggesting that parameter estimates are trustworthy. Estimates for consistency and occasion specificity showed that most items and scales of the heiQ™ are primarily influenced by person factors rather than by measurement situations. In addition, the heiQ™ scales show (very) good reliability coefficients. The TCO coefficients are comparable with or even higher than those of other measures of situation-unspecific mental health [11, 14]. For example, it was shown that some measures of depression show OSpe > 0.4, indicating that the scales are affected to a large extend by the measurement situation [46]. These results demonstrate that the heiQ™ fulfills important requirements to be a valid instrument to evaluate stable effects of self-management programs.

However, the results for *Health-directed activity* must be interpreted with caution. The model could only be estimated by fixing the residual variance of $\text{Trait}_{2\text{-}1}$ to 0.01; however, this procedure assumes that the population value of this parameter is close to 0 and that the improper solution is not caused by wrong model specification [45, 47].

As expected, most items of the heiQ™ showed lower reliability values than their respective scales, supporting the notion that researcher should use scales rather than individual items. Interestingly, some items show higher UCO values than CCO values, indicating that most part of the stable variance measured by these items are not shared by the reference item. For example, in *Self-monitoring and insight*, the CCO values of item 11 ("I have a very good

Table 5 Means and standard deviations of latent trait variables ($Trait_1$, $Trait_2$) and the latent difference variable ($Trait_{2\text{-}1}$), effect sizes and intercepts of latent state residual variable on measurement occasion 2 (SR_2)

heiQ^{TM} scale	$Trait_1$ (SD)	$Trait_2$ (SD)	$Trait_{2\text{-}1}$ (SD)	$ES_{Trait2\text{-}1}$ (SE)	I_{SR2} (SE)	$ES_{Obs3\text{-}1}$ (SE)
Positive and active engagement in life	2.69 (0.31)	2.84 (0.51)	0.14* (0.19)	0.26* (0.05)	0.12* (0.03)	0.17* (0.04)
Health-directed activities	2.61 (0.57)	2.86 (0.42)	0.24* (0.14)	0.43* (0.05)	0.22* (0.03)	0.28* (0.04)
Skill and technique acquisition	2.96 (0.46)	3.22 (0.43)	0.26* (0.23)	0.57* (0.06)	0.05* (0.02)	0.49* (0.04)
Constrictive attitudes and approaches	3.18 (0.52)	3.22 (0.50)	0.04 (0.17)	0.08 (0.05)	0.07* (0.02)	0.07* (0.04)
Self-monitoring and insight	2.92 (0.51)	3.21 (0.43)	0.29* (0.31)	0.57* (0.05)	0.03 (0.02)	0.45* (0.03)
Health service navigation	3.28 (0.47)	3.35 (0.45)	0.07* (0.08)	0.14* (0.04)	0.00 (0.02)	0.14* (0.03)
Social integration and support	2.91 (0.65)	2.95 (0.57)	0.04 (0.17)	0.07 (0.04)	0.09* (0.02)	0.06 (0.03)
Emotional distress	2.28 (0.78)	1.92 (0.78)	-0.36* (0.26)	-0.46* (0.04)	0.02 (0.02)	-0.42* (0.03)

SD: standard deviation; SE: standard error; I_{sr2}: intercept of SR_2 on $Trait_2$ (see text); $ES_{Trait2\text{-}1}$: differences between latent traits in effect size metric

($M(Trait_{2\text{-}1})/SD(Trait_1)$); $ES_{Obs3\text{-}1}$: effect size of manifest heiQ^{TM} scales between T3 and T1 (see text)

understanding of when and why I am supposed to take my medication") are somewhat low (CCO = 0.12–0.14), while the UCO values are comparatively high (UCO = 0.28–0.31). One might conclude that item 11 is a poor indicator of the measured construct. However, another explanation might be that the construct measured by this scale is not fully captured by the reference item (in this case, item 16 ["When I have health problems, I have a clear understanding of what I need to do to control them")]. Note that in a CT − C(M− 1) model the contents of the state- and trait variables are determined by respective reference items.

Schwarze and Spanier et al. (in prep) demonstrated short-term changes in the heiQ^{TM} scales after inpatient rehabilitation. Our analyses complement their results and showed that changes in self-management skills (*Skill and technique acquisition*, *Self-monitoring and insight*) or emotional well-being (*Emotional distress*, *Active engagement in life*) of medium size were stable at least over a period of 3 months. These results are also in line with the assumption of a trait-change in the constructs measured by the heiQ^{TM} scales. However, it needs to be noted that we did not (and could not) prove that the trait-change model is correct; we only showed that the model provides a plausible interpretation of the data.

Also, in line with Schwarze and Spanier et al. (in prep), changes in most heiQ^{TM}-scales can be found in all diagnostic groups. This result supports the generic definition of the heiQ^{TM} scales, as all self-management programs pursue similar goals (e.g., enhancing self-management). Nevertheless, in many scales, $Trait_1$ and $Trait_{2\text{-}1}$ are clearly affected by type of chronic disease. This may reflect differences in clinical presentation, course of diseases, and demands of treatment among these chronic conditions. For example, the lower values of patients with rheumatic diseases on $Trait_1$ and $Trait_{2\text{-}1}$ in some heiQ^{TM}-scales may reflect that rheumatic diseases have high impact on health status and that both symptoms and disease trajectories are difficult to control. On the other hand, the higher gains of asthma patients in *Skill and technique acquisition* may reflect that training to control asthma attacks is a major part of the treatment [48].

All heiQ^{TM}-scales show scalar invariance over time according to the latent state factor. In contrast, Nolte et al. [49] found some heiQ^{TM} items to be scalar invariant; however, they used an earlier (English) version of the heiQ^{TM} (42, item, 6-point response scale) and a different (and probably stricter) criterion to identify non-invariant items, i.e., the chi^2-difference test [50]. In our study, non-invariance is defined by the size of the difference in a parameter over time [33], i.e., EPC > 0.25.

Nearly all heiQ^{TM} scales show non-invariant factor loadings in some items according to method factors. However, since the method factors in the CT − C(M − 1) approach are defined as residuals with respect to the latent trait factors/latent state residual factors and have an expected mean of 0, non-invariant factor loadings of method factors do not change the meaning of the latent trait factors/latent state residual factors nor do they affect latent means.

Limitations

The main limitation of this study is that some model assumptions could not be tested empirically. For example, whether mean differences in heiQ^{TM} items or scales should be modeled as trait-changes or as state-changes is a conceptual decision, not an empirical result. Another challenging problem is that our LST models could only be

Table 6 Impact of disease group, age and sex on latent trait variables ($Trait_1$) and latent difference variables ($Trait_{2\text{-}1}$) of each heiQ™ scale

Predictor	$Trait_1$	IBD		Orthopedic		Asthma		Rheumatism		Age		Sex	
heiQ™ scale	$b_{Trait2\text{-}1}$ (SE)	b_{Trait1} (SE)	$b_{Trait2\text{-}1}$ (SE)	b_{Trait1} (SE)	$b_{Trait2\text{-}1}$ (SE)	b_{Trait1} (SE)	$b_{Trait2\text{-}1}$ (SE)	b_{Trait1} (SE)	$b_{Trait2\text{-}1}$ (SE)	b_{Trait1} (SE)	$b_{Trait2\text{-}1}$ (SE)	b_{Trait1} (SE)	$b_{Trait2\text{-}1}$ (SE)
Positive and active engagement in life	−0.14* (0.07)	0.07 (0.07)	−0.03 (0.05)	0.07 (0.06)	0.02 (0.04)	$<\lvert -0.01 \rvert$ (0.05)	0.08* (0.03)	−0.12 (0.07)	−0.10* (0.05)	0.004 (0.003)	−0.004 (0.002)	−0.21* (0.07)	0.10* (0.04)
Health-directed activities[a]	−0.27* (0.04)	0.06 (0.07)	−0.07 (0.05)	−0.07 (0.07)	0.05 (0.04)	−0.07 (0.05)	0.08* (0.03)	0.12 (0.07)	−0.10* (0.04)	0.013* (0.004)	−0.007* (0.003)	0.06 (0.07)	0.05 (0.04)
Skill and technique acquisition	−0.23* (0.09)	0.08 (0.06)	−0.03 (0.05)	−0.05 (0.05)	−0.04 (0.05)	0.23* (0.04)	0.13* (0.06)	−0.21* (0.05)	−0.04 (0.05)	0.007* (0.003)	−0.004* (0.002)	−0.05 (0.05)	0.05 (0.04)
Constructive attitudes and approaches	−0.07 (0.06)	−0.11 (0.07)	0.04 (0.04)	0.01 (0.05)	−0.03 (0.04)	0.14* (0.04)	0.03 (0.03)	−0.12* (0.06)	−0.04 (0.04)	0.006 (0.003)	−0.002 (0.02)	−0.04 (0.06)	0.11 (0.04)
Self-monitoring and insight	−0.36* (0.06)	0.01 (0.06)	−0.06 (0.04)	−0.03 (0.05)	−0.07* (0.03)	0.17* (0.04)	0.15* (0.03)	−0.19* (0.06)	−0.08* (0.04)	0.007* (0.003)	−0.002 (0.002)	−0.01 (0.06)	0.06 (0.04)
Health service navigation	−0.03 (0.07)	−0.06 (0.05)	0.07 (0.04)	0.08 (0.05)	−0.13 (0.04)	0.12* (0.04)	0.01 (0.03)	−0.20* (0.06)	0.06 (0.04)	<0.001 (0.003)	0.001 (0.002)	−0.05 (0.05)	0.04 (0.04)
Social integration and support	−0.15* (0.04)	0.05 (0.08)	0.02 (0.04)	0.14* (0.06)	−0.05 (0.03)	0.07 (0.05)	0.04 (0.05)	−0.27* (0.07)	−0.03 (0.04)	0.010* (0.004)	−0.003 (0.002)	0.04 (0.07)	0.05 (0.04)
Emotional distress	−0.04 (0.05)	−0.04 (0.08)	0.07 (0.05)	−0.20* (0.07)	−0.04 (0.05)	−0.23* (0.06)	−0.07 (0.04)	0.55* (0.08)	0.07 (0.06)	−0.011* (0.004)	0.007* (0.003)	0.23* (0.07)	−0.15* (0.05)

b: unstandardized regression coefficient; SE: standard error; IBD: inflammatory bowel disease

* Statistical significant coefficient

[a] Residual variance of $Trait_{2\text{-}1}$ restricted to 0.01

identified by restricting the latent state residual variances to be stable over time. Neither the correctness nor the consequences of this assumption on estimates of latent trait–state variances can be tested. To identify a latent change model without such restrictions, a further measurement occasion for trait1 is necessary [13, 14]. Unfortunately, data from two or more time points before an intervention are often unavailable or difficult to obtain.

Using the CT − C(M − 1) approach, parts of our results (e.g., CCO/UCO coefficients, latent trait means) strongly depend on the choice of an appropriate reference indicator. However, small CCO values in some items raise the questions whether the main content of all heiQ™ scales can be captured by a single item, whether the best items were chosen as reference indicator or whether items with low CCO values should be deleted from the scale. Further psychometric studies of the heiQ™ may clarify these issues.

Though we included disease group as predictor of $Trait_1$ and $Trait_{2\text{-}1}$, we did not examine whether LST parameters differed between disease groups since the numbers of patients in the subgroups would have been too small to yield sufficiently robust estimates.

Conclusion

The heiQ™ can be used with confidence in a variety of settings and conditions. Our results indicate that it may be a useful tool for assessing stable effects in important outcomes of self-management programs over time, e.g., changes in self-management skills or emotional well-being. Furthermore, we showed that applying LST theory can give further insights into the psychometric properties of measurement instruments in the health sciences.

Acknowledgments The authors wish to thank our cooperation clinics: Rehabilitation Center Bad Eilsen, Hospital Bad Bramstedt, Hospital Bad Oexen, Hospital Bad Reichenhall, Hospital Norderney, Deegenberg Hospital Bad Kissingen and Rehabilitation Center Bad Mergentheim Hospital Taubertal. This project was funded by the German Federal Ministry of Education and Research (Bundesministerium fuer Bildung und Forschung). Professor Osborne was supported in part by an Australian National Health and Medical Research Council Population Health Career Development Award (#400391).

Conflict of interest The authors state that there are no conflicts of interests.

References

1. Lorig, K. R. (2003). Self-management education: More than a nice extra. *Medical Care, 41*(6), 699–701.

2. Lorig, K. R., & Holman, H. (2003). Self-management education: History, definition, outcomes, and mechanisms. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine, 26*(1), 1–7.
3. Osborne, R. H., Spinks, J. M., & Wicks, I. P. (2004). Patient education and self-management programs in arthritis. *Medical Journal of Australia, 180*(5 Suppl), S23–S26.
4. Faller, H., Reusch, A., & Meng, K. (2011). DGRW-Update: Patientenschulung. *Rehabilitation, 50*(5), 284–291.
5. Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state–trait theory and research in personality and individual differences. *European Journal of Personality, 13*(5, Spec Issue), 389–408.
6. Nesselroade, J. R. (2004). Intraindividual variability and short-term change. *Gerontology, 50*(1), 44–47.
7. Eid, M. (1995). *Modelle der Messung von Personen in Situationen*. Weinheim: Psychologie Verlags Union.
8. Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment, 8*(2), 79–98.
9. Steyer, R., Geiser, C., & Fiege, C. (2012). Latent state–trait models. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology* (Vol. 3, pp. 291–308)., Data analysis and research publication Washington, DC: American Psychological Association.
10. Geiser, C., & Lockhart, G. (2012). A comparison of four approaches to account for method effects in latent state–trait analyses. *Psychological Methods, 17*(2), 255–283.
11. Courvoisier, D. S., Agoritsas, T., Glauser, J., Michaud, K., Wolfe, F., Cantoni, E., et al. (2012). Pain as an important predictor of psychosocial health in patients with rheumatoid arthritis. *Arthritis Care & Research, 64*(2), 190–196.
12. Courvoisier, D. S., Eid, M., & Nussbeck, F. W. (2007). Mixture distribution latent state–trait analysis: Basic ideas and applications. *Psychological Methods, 12*(1), 80–104.
13. Eid, M., & Hoffmann, L. (1998). Measuring variability and change with an item response model for polytomous variables. *Journal of Educational and Behavioral Statistics, 23*(3), 193–215.
14. Steyer, R., Krambeer, S., & Hannöver, W. (2004). Modeling latent trait-change. In K. Van Montfort, H. Oud, & A. Satorra (Eds.), *Recent developments on structural equation modeling: Theory and applications* (pp. 337–357). Amsterdam: Kluwer Academic Press.
15. Osborne, R. H., Elsworth, G. R., & Whitfield, K. (2007). The Health Education Impact Questionnaire (heiQ): An outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. *Patient Education and Counseling, 66*(2), 192–201.
16. Osborne, R. H., Batterham, R., & Livingston, J. (2011). The evaluation of chronic disease self-management support across settings: The international experience of the health education impact questionnaire quality monitoring system. *The Nursing clinics of North America, 46*(3), 255–270.
17. Epstein, J., Osborne, R. H., Elsworth, G. R., Beaton, D. E., & Guillemin, F. (2013). Cross-cultural adaptation of the Health Education Impact Questionnaire: Experimental study showed expert committee, not back-translation, added value. *Journal of Clinical Epidemiology*. doi: 10.1016/j.jclinepi.2013.07.013.
18. Schuler, M., Musekamp, G., Faller, H., Ehlebracht-Konig, I., Gutenbrunner, C., Kirchhof, R., et al. (2013). Assessment of proximal outcomes of self-management programs: Translation and psychometric evaluation of a German version of the Health Education Impact Questionnaire (heiQ). *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 22*(6), 1391–1403.
19. Ehlebracht-König, I., & Bönisch, A. (2007). Beispiel einer qualitätsgesicherten Schulung bei Patienten mit Spondylitis ankylosans. *Praxis Klinische Verhaltensmedizin und Rehabilitation, 75*, 15–20.
20. Ehlebracht-König, I., Bönisch, A., & Pönicke, J. (2009). Fraktionierte rehabilitation: Ergebnisse einer randomisierten, kontrollierten Studie. *Rehabilitation, 48*, 30–38.
21. Schultz, K., Petro, W., Müller, C., & Schwiersch, M. (2000). Asthma-Verhaltenstraining mit Erwachsenen: Konzepte und Materialien. In F. Petermann & P. Warschburger (Eds.), *Asthma bronchiale* (pp. 275–294). Göttingen: Hogrefe.
22. Schultz, K., Schwiersch, M., Petro, W., Mühlig, S., & Petermann, F. (2000). Individualisiertes, modular strukturiertes Patientenverhaltenstraining bei obstruktiven Atemwegserkrankungen in der stationären Rehabilitation. *Pneumologie, 54*, 296–305.
23. Deutsche Rentenversicherung Bund. (2011). *Reha Therapiestandards Chronischer Rückenschmerz*. Berlin: Deutsche Rentenversicherung Bund.
24. Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives, 4*(1), 10–18.
25. Jelitte, M., & Schuler, M. (2012). Messen wir immer das Gleiche? Zur Invarianz von Messungen und Response-Shift in der Rehabilitation - Teil 2. *Die Rehabilitation, 51*(6), 415–423.
26. Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 203–240). Washington, DC: American Psychological Association.
27. Schuler, M., & Jelitte, M. (2012). Messen wir bei allen Personen das Gleiche? Zur Invarianz von Messungen und Response Shift in der Rehabilitation - Teil 1. *Die Rehabilitation, 51*(5), 332–339.
28. Geiser, C., Eid, M., West, S. G., Lischetzke, T., & Nussbeck, F. W. (2012). A comparison of method effects in two confirmatory factor models for structurally different methods. *Structural Equation Modeling, 19*(3), 409–436.
29. Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika, 65*(2), 241–261.
30. Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study. *Structural Equation Modeling, 14*(3), 453–463.
31. Oberski, D. J. (2009). *Jrule for Mplus (Version 0.91)*. http://wiki.github.com/daob/JruleMplus/.
32. Steinmetz, H. (2013). Analyzing observed composite differences across groups is partial measurement invariance enough? *Methodology-European Journal of Research Methods for the Behavioral and Social Sciences, 9*(1), 1–12.
33. Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling, 16*(4), 561–582.
34. van der Veld, W. M., & Saris, W. E. (2011). Causes of generalized social trust. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *European association for methodology series* (pp. 207–247). New York, NY: Routledge/Taylor & Francis Group.
35. Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures—The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456–466.
36. Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research, 2*(1), 21–33.
37. Leonhart, R. (2004). Effektgrößenberechnung bei Interventionsstudien. *Die Rehabilitation, 43*(4), 241–246.
38. Kazis, L. E., Anderson, J. J., & Meenan, R. F. (1989). Effect sizes for interpreting changes in health status. *Medical Care, 27*(3 Suppl), S178–S189.

39. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2., 2 print. ed.). Hillsdale, NJ u.a.: Erlbaum.
40. West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality, 64*(1), 1–48.
41. Eid, M., & Diener, E. (2004). Global judgments of subjective well-being: Situational variability and long-term stability. *Social Indicators Research, 65*(3), 245–277.
42. Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology, 76*(4), 662–676.
43. Muthén, L. K., & Muthén, B. (2010). *Mplus user;s guide*. Los Angeles: Muthén & Muthén.
44. Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.
45. Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research, 29*(4), 468–508.
46. Mohiyeddini, C., Hautzinger, M., & Bauer, S. (2002). A latent state–trait analysis on assessing trait and state components of three instruments for measuring depression: ADS, BDI, and SDS. *Diagnostica, 48*(1), 12–18.
47. Kolenikov, S., & Bollen, K. A. (2012). Testing negative error variances: Is a Heywood case a symptom of misspecification? *Sociological Methods & Research, 41*(1), 124–167.
48. Muhlig, S., Schultz, K., de Vries, U., & Petermann, F. (2000). Grundlagen der Patientenschulung bei Asthma. In F. Petermann & P. Warschburger (Eds.), *Asthma bronchiale* (pp. Seiten 147–174). Gottingen: Hogrefe.
49. Nolte, S., Elsworth, G. R., Sinclair, A. J., & Osborne, R. H. (2009). Tests of measurement invariance failed to support the application of the "then-test". *Journal of Clinical Epidemiology, 62*(11), 1173–1180.
50. Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research, 14*(3), 587–598.