Psychological Test and Assessment Modeling, Volume 52, 2010 (1), 29-47

# Direct measurement of psychotherapeutic outcomes: Experimental construction and validation of a brief scale

*Günter Krampen*[1]

**Abstract**

Results of two studies on the construction of a brief questionnaire for the direct measurement of psychotherapeutic outcomes are presented. Test construction follows the theory of change in integrative, differential psychotherapy. Items focus on changes in behavior and experience with reference to increases in self-efficacy and improved coping in patients. Study 1 included 150 outpatients of 14 psychotherapists. Controlling for diagnosis, gender, and age, patients were randomized to a therapy group ($n$ = 75) or a waiting-list control group ($n$ = 75). After three months of waiting or psychotherapy, all subjects completed the "Questionnaire of Personal Changes" consisting of 12 temporal comparatively formulated items. Group comparisons indicate the change sensitivity of the items and the scale score for psychotherapeutic effects. Study 2 followed a naturalistic design including 275 psychotherapy outpatients who completed the questionnaire 14-20 weeks after start of therapy. Results confirm acceptable psychometric properties of the scale with reference to item parameters and reliability. Convergent validity of the scale score is empirically supported by significant correlations to clinically relevant indicators of psychotherapeutic outcomes from indirect measurements of change.

Key words: direct measurement of change, psychotherapeutic outcomes, treatment effectiveness evaluation, psychometrics, psychological assessment

[1] *Correspondence concerning this article should be addressed to:* Prof. Günter Krampen, PhD, University of Trier, Department of Psychology, 54286 Trier, Germany; email: krampen@uni-trier.de

## Introduction

In theory, the approach of direct measurement of change has been discussed for a number of decades as a complementary strategy for approximating the "true" changes that patients with mental disorders experience during the course of their psychotherapy. However, in reality, psychometric constructions of direct measures of change have continued to be rather reserved and are restricted to instruments developed *ad hoc*. Therefore, in the present contribution, an empirically well-founded measure based upon the theory of change in integrative, differential psychotherapy is presented. Items in the measure were designed to focus on changes in both the behavior and experiences of patients involved in psychotherapy while referring to their increases in self-efficacy and improved coping. After the initial empirical tests of the experimental construction of this scale for the direct measurement of psychotherapeutic outcomes in adult outpatients, the central aspects of the scale's reliability, validity, and change sensitivity are empirically tested in a naturalistic study with an independent sample of psychotherapy outpatients.

More than 40 years ago Bereiter (1963, p. 3) entitled a conference proceedings' contribution "Some persisting dilemmas in the measurement of change". In this article he critically and constructively described three problems in measuring changes. Since outlining these problems, much work has been done to correct them. The first problem is known as the "over-correction-under-correction dilemma" which refers to the regression to the mean--and was interpreted by Bereiter as the difficulty to correct for unreliability of pretest scores (see Hsu, 1995; Jacobson & Truax, 1991; see also Cribbie & Jamieson, 2000; Foerster, 1995; Jamieson, 1995; Johnson, Dow, Lynch, & Hermann, 2006; Tan, Imbos, Does, & Theunissen, 1995). Researchers have been busy working on the second problem as well--the "unreliability-invalidity dilemma"--that Bereiter (1963, p. 20) had stated to be a "false one" because "the meaningfulness of change scores does not depend on a test's measuring 'the same thing' on two occasions" (e.g., Beutler & Hamblin, 1986; Embretson, 2006; Jacobson & Truax, 1991; May & Hittner, 2003; Reise & Haviland, 2005; Sandell, 1987; Willett, 1989; Zielke, 1980; for an overview see, e.g., Lambert & Hill, 1994; Nesselroade & Ghisletta, 2003). However, the third problem--the "physicalism-subjectivism dilemma"--which Bereiter (1963, p. 20) concluded to be "the only true one" has resulted in, at least up to now, rather weak methodological attempts to construct meaningful direct measures of change.

Thus, in the past 40 years, psychometric constructions of direct measures of change have continued to be rather reserved and are restricted to *ad hoc* developed instruments, which are for the most part constructed on the foundation of pre-experimental designs (e.g., Ard & Cook, 1977; Balzer, 1989; Miller, 1992; Mintz, Luborsky, & Christoph, 1979; Schaffer, Murillo, & Michael, 1980; Weinstock & Meier, 2003) and even rarely on the foundation of quasi-experimental designs (e.g., Vermeersch, 1998). Thus, the potentials of "change items" and direct change measures--Bereiter (1963) even put forth suggestions for the improvement of change measurements by their complimentary application--are rarely taken into account seriously even now.

In accordance with Bereiter's concept of "traits of change", the "way is open for a whole new order of hypothetical constructs" (Bereiter, 1963, p. 16) for the development of meaningful measures of subjective changes. However, in present-day psychotherapy, we are in need of intervention-based item selection procedures: Item selection (and scale scores) must refer not only to standard item and test analyses but also to indicators of item and scale score sensitivity to change in response to interventions. Prerequisites are true experimental designs with randomization (RCTs) of individuals (patients) to a treatment and a waiting-list control group. Item and scale score sensitivity for change should be computed by comparisons of specific changes in the treatment group following psychological intervention and the waiting-list control group without intervention.

Results of the experimental construction of the Questionnaire of Personal Changes (Q-PC), a newly created scale for the direct measurement of psychotherapeutic outcomes in adult outpatients are presented in Study 1. Scale construction was based on the foundations and theory of change in integrative and differential psychotherapy (e.g., Grawe, 2004). Items focus on subjectively perceived changes in behavior (six items: I can relax much better; I can unwind better and take it easy; I sleep better; I take less medication; I have more stamina and do not give up as easily; I can concentrate much better; see Table 2) and experience (six items: I feel less anxious thinking about the future; Overall I feel healthier; I feel calmer and more well-balanced; I feel better; I cope with unexpected events more easily; I deal with stress and pressure better; see Table 2). The items were constructed with reference to the general psychotherapeutic effect. The items focus the outcome factor "active help in problem solving and symptom reduction" which refers to improvements in self-efficacy and in coping with life problems in psychotherapy patients (i.e., mastery-oriented interventions in a psychotherapeutic problem-solving perspective; Grawe, 2004). Item formulations are temporally comparative to the individual start of psychotherapy. Instructions of the "Questionnaire of Personal Changes" (Q-PC) for patients are formulated as follows: "Please think back to the time prior to beginning your treatment. …. The treatment started on _____ (please enter the date!), approximately ____ weeks ago (please enter the number of weeks!). Try to remember what you did at that time and how you felt. For the following statements, please indicate the kind of changes you experienced, in one direction or the other, during your treatment. Use the scale provided below to evaluate these changes: +3 = strong positive change, +2 = medium positive change, +1 = weak positive change, 0 = no change, -1 = weak negative change, -2 = medium negative change, -3 = strong negative change" (see Appendix).

Psychometric quality of the brief scale is empirically cross-analyzed in Study 2. Data is presented on the statistical item parameters, scale score reliability, item and scale score sensitivity for changes in response to psychotherapy. In addition, selected aspects of its convergent and discriminative validity and the scale's sensitivity for response sets are analyzed. Study 2 follows a naturalistic design including a larger sample of psychotherapy outpatients. Convergent validation refers to hypotheses on significant correlations of the Q-PC to indirect pretest-posttest-difference measures of general complaints and symptoms (replication analysis of results of Study 1) as well as neuroticism. Furthermore, the hypothesis is tested that these convergent validity correlation coefficients are significantly stronger than the discriminative validity correlations referring to extraver-

sion, openness to experience, agreeablessness, conscientiousness, cognitive speed, and short-term memory.

## Study 1

Study 1 is an empirically rigorous attempt to construct a brief scale for the direct measurement of psychotherapeutic outcomes that is oriented towards the theory of change in integrative, differential psychotherapy. Items focus on changes in behavior and experience during the course of psychotherapy with reference to the patient's increases in self-efficacy and improved coping. Of particular focus are the central aspects of the scale's reliability, the change sensitivity of its items and its scores as well as its convergent validity to indirect measures of psychotherapeutic outcomes. In addition, special attention is given to the scale's correlations with response sets.

### Method

*Participants.* Participants were an unselected sample of 150 adult patients who consulted a total of 14 psychotherapists in private practice for outpatient psychotherapy because of dominant symptoms of depression, anxiety, adjustment/stress, or somatoform disorders (eligibility criteria). Thus, sampling refers to the consecutive admissions of all patients (over a period of three years) with these dominant symptoms. These are the most frequent reasons for seeking psychotherapy treatment in the geographical region in which the study took place. Patients were referred to psychotherapy by physicians ($n$ = 72), other psychotherapists ($n$ = 44), or sought treatment themselves ($n$ = 34). Physical examinations excluded medical factors. Therapy costs were covered by the patients' health insurance.

Age of patients ranged from 18 to 72 years ($M$ = 33.8, $SD$ = 11.7). There are 84 females and 66 males in the sample. Occupational status and level of education indicate that the majority of the patients were middle class.

After being informed about the treatment process in personal interviews, patients provided their individual consent to psychotherapy and to the diagnostic and evaluation procedures by signing informed consent forms with their full names. Moreover, patients were guaranteed that their records (for study purposes) would be recorded anonymously and that later rescinding of consent would have no effect on the treatment. Throughout the course of the study, there were neither treatment refusals nor treatment dropouts.

Initial stage clinical interviews according to DSM-IV (SCID-I and SCID-II; First, Gibbon, Spitzer, & Williams, 1996; First, Spitzer, Gibbon, & Williams, 1996) were conducted with all patients. Primary diagnoses were depression disorders (DSM-IV: 296.xx, $n$ = 64), anxiety disorders (DSM-IV: 300.2x and 300.02, $n$ = 38), adjustment and stress disorders (DSM-IV: 309.xx, $n$ = 30), and somatoform disorders (DSM-IV: 308.x, $n$ = 18).

*Psychotherapists.* Psychotherapy sessions involving only the patient and psychotherapist were conducted by 14 experienced psychotherapists in private practice (job experience: 14-31 years) and were covered by the patients' health insurance. All had professional licensing and full certifications in cognitive-behavioral psychotherapy ($n$ = 12), psychodynamic therapy ($n$ = 4), client-centered psychotherapy ($n$ = 6), and/or relaxation therapy ($n$ = 14). Their basic therapeutic orientation refers to the general psychological therapy approach (Grawe, 2004) focusing on individualized and adaptive treatment, which was professionally supervised on a regular basis (at least once every two weeks).

*Procedure.* A randomized design was employed. After pretest and randomization (controlling for age, gender, and mental disorder), patients in Group I ($n$ = 75) participated in outpatient psychotherapy for 12 weeks with one session weekly. Patients in Group II ($n$ = 75) were the waiting-list controls for 12 weeks (without psychotherapy) who began their psychotherapy after waiting 12 weeks. Randomization controlling for age, gender, and mental disorder of patients succeeded in comparable groups (see Table 1). This was extended by *post hoc* comparisons of the average depression scores (Beck Depression Inventory; BDI) and global symptom complaints score (A-SYM; see below) as well as the time-span under direct measurement of change confirming parallel groups in these pre-test variables (see Table 1).

*Measures.* Initial stage clinical interviews according to DSM-IV were conducted by the therapists with SCID-I and SCID-II (First et al., 1996) including the *Global Assessment of Functioning* (GAF). GAF was repeated by therapists after 12 weeks of psychotherapy versus waiting. In addition pre- and post-tests were conducted with the German versions of the *Beck Depression Inventory* (BDI; Beck, Hautzinger, Bailer, Worall & Keller, 1995; split-half reliability in the present sample: $r_{tt} > .83$), the *Hopelessness-Scale* (H-Scale; Beck, Weissmann, Lester & Trexler, 1974; Krampen, 1994; $r_{tt} > .91$), and a *German Symptom Checklist* (A-SYM; Krampen, 2007) consisting of eight items per scale to measure physical and mental stress (item examples: exhaustion and tiredness, feeling of unease, physical tension; $r_{tt} > .78$), nervousness and inner strain (irritability, mental tension, tearfulness; $r_{tt} > .80$), psychophysiological deregulation (vertigo, heart pounding, poor digestion; $r_{tt} > .75$), behavior and achievement disorders (concentration problems, problems in social interaction, forgetfulness; $r_{tt} > .79$), pain distress (headache, rheumatic pain, earache; $r_{tt} > .77$), problems in self-control (indecisiveness, shyness, self-distrust; $r_{tt} > .82$) as well as global symptom complaints (all 48 items; $r_{tt} > .94$) during the last two weeks.

At post-test, data with empirically proven scales on response sets like social desirability *(Marlowe-Crowne Desirability Scale, MCDS;* Crowne & Marlowe, 1964; $r_{tt} = .86$), faking (L-Scale of the *Eysenck Personality Inventory, EPI;* Eysenck, 1970; $r_{tt} = .71$), and openness in questionnaire responses (O-Scale of the *Freiburger Persönlichkeitsinventar, FPI-R;* Fahrenberg, Hempel & Selg, 2001; $r_{tt} = .76$) were gathered as well.

All participants responded to the "Questionnaire of Personal Changes" (Q-PC; see Appendix) either 12 weeks after psychotherapy (12 sessions) versus after 12 weeks of waiting. Psychotherapy patients were asked, at post-test, to answer the comparatively formu-

**Table 1:**
Variables Controlled for in Randomization and Parallelization Variables as well as Direct Measure of Change Scale (Q-PC) Statistics in 75 Psychotherapy Outpatients and 75 Waiting-List Controls with Group Comparisons (Study 1)

| Variable | Psychotherapy Patients | | | Waiting-List Controls | | | *t-Test* (*df* =148) |
|---|---|---|---|---|---|---|---|
| | *f* | *M* | *SD* | *f* | *M* | *SD* | |
| Age | - | 33.5 | 10.91 | - | 34.1 | 12.03 | 0.78 |
| Females | 42 | - | - | 42 | - | - | |
| Males | 33 | - | - | 33 | - | - | |
| Diagnosis | | | | | | | |
| Major Depression (DSM-IV: 296.xx) | 32 | - | - | 32 | - | - | |
| Anxiety Disorders (DSM-IV: 300.2x, 300.02) | 19 | - | - | 19 | - | - | |
| Adjustment/Stress Disorders (DSM-IV: 309.xx) | 15 | - | - | 15 | - | - | |
| Somatoform Disorders (DSM-IV: 300.8x) | 9 | - | - | 9 | - | - | |
| BDI: Depression | - | 23.4 | 9.93 | - | 24.2 | 10.17 | 1.21 |
| A-SYM: Complaints and Symptoms | - | 70.8 | 12.25 | - | 71.2 | 13.54 | 1.47 |
| Direct Measure of Change | $r_{tt}$ | | | $r_{rr}$ | | | |
| Change Rating Duration (weeks) | - | 12.7 | 2.04 | - | 12.3 | 1.87 | 0.55 |
| Direct Measure of Change Scale (Q-PC) | .94 | +15.2 | 11.8 | .91 | -4.3 | 12.1 | 11.53** |

**$p < .01$

lated items for changes since start of treatment (12 weeks before), waiting-list controls were asked to answer the same items for perceived changes since their scheduling of treatment and the initial clinical interview (pre-test 12 weeks before) without actually attending any psychotherapy sessions during that time.

## Results

Means and standard deviations of the brief scale's items on personal changes are summarized for both groups in Table 2. Both the absolute comparisons of item means (with the positive vs. negative signs indicating more or less improvements vs. deterioration, respectively) and the statistical mean comparisons point to significant differences between the psychotherapy group and the waiting-list control group. For all of the 12 items there are positive ratings of change in the psychotherapy patients, which are statistically significant different from the, on average, negative ratings of change in the waiting-list controls. Effect sizes are large in terms of Cohen's (1988) guidelines. These results confirm--in a true experimental design with randomization--the sensitivity of the items to specific changes following psychotherapy in contrast to no changes or negative changes following only waiting time without intervention.

The items sum up to a scale score sensitive to specific changes after psychotherapy: Mean scale score is $M$ = +15.2 in the psychotherapy group, which indicates distinct improvements during psychotherapy and $M$ = -4.3 in the waiting-list controls, indicating some deterioration during waiting time. The mean difference is statistically significant and shows a large effect size ($d$ = 1.64). In addition, item-scale correlations exceed $r_{it} \geq .46$ and reliability coefficients are comparably high for both groups (see Table 2).

Intercorrelations between the direct measure of change scale (Q-PC) score and the indirect measures of change (difference scores) are summarized in the upper part of Table 3. All of them are statistically significant, with common variances between 7 % (Pain Distress) and 27 % (GAF). In summary, we can conclude that indirect measures of change resulting from therapists' global functioning ratings (DSM-IV Axis 5: GAF) and of the psychometric measures of depression, hopelessness, and mental stress and strains show stronger correlations with the direct measure of change (mean common variance: 18%), while those indirect measures of change resulting from patients' self-report data on somewhat more physical symptoms (psychophysiological deregulation, pain distress) show somewhat weaker correlations with the direct measure of change (9 % mean common variance).

In comparison, the correlations of the direct measure of change scale with the response set scales' scores are rather low. Faking (Lying Scale) and openness in questionnaire responding (O-Scale) are not significantly correlated, the correlation with social desirability (MCDS) is statistically significant, albeit low (common variance: 5.7 %) with the direct measure of personal changes (see Table 4).

**Table 2:**

Means, Standard Deviations, Mean Comparisons, and Effect Sizes of Items and Scale Score in the Psychotherapy Group (n = 75) and the in Waiting-List Control Group (n = 75; Study 1) as well as Item Statistics und Scale Statistics in a Sample of 275 Psychotherapy Outpatients (Study 2)

| Item-No. | Item | Therapy Patients | | Control Group | | *t-Test* | | 275 Psychotherapy Patients | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *SD* | (*df* = 148) | *d* | *M* | *SD* | $r_{it}$ |
| 01. | I can relax much better. | +1.4 | 1.14 | -0.2 | 1.18 | 6.41** | 1.38 | +1.5 | 1.21 | .61 |
| 02. | I can unwind better and take it easy. | +1.3 | 1.21 | -0.4 | 1.20 | 5.89** | 1.41 | +1.2 | 1.29 | .65 |
| 03. | Overall I feel healthier. | +1.3 | 1.40 | -0.2 | 1.37 | 3.14** | 1.08 | +1.2 | 1.38 | .78 |
| 04. | I feel less anxious thinking about the future. | +1.4 | 1.43 | -0.8 | 1.39 | 7.31** | 1.56 | +1.5 | 1.42 | .72 |
| 05. | I feel calmer and more well-balanced. | +1.2 | 1.37 | -1.1 | 1.41 | 6.83** | 1.65 | +1.3 | 1.39 | .82 |
| 06. | I sleep better. | +1.1 | 1.52 | -0.3 | 1.55 | 5.58** | 0.91 | +1.0 | 1.64 | .46 |
| 07. | I take less medication. | +0.8 | 1.22 | 0.0 | 1.31 | 2.32* | 0.63 | +0.7 | 1.33 | .47 |
| 08. | I have more stamina and do not give up as easy. | +1.4 | 0.96 | -0.5 | 1.13 | 4.88** | 1.82 | +1.4 | 1.01 | .44 |
| 09. | I can concentrate much better. | +1.3 | 1.28 | -0.1 | 1.37 | 3.87** | 1.05 | +1.4 | 1.30 | .45 |
| 10. | I cope with unexpected events more easily. | +1.1 | 1.23 | -0.2 | 1.19 | 2.59** | 1.07 | +1.3 | 1.18 | .68 |
| 11. | I feel better. | +1.6 | 1.42 | -0.3 | 1.35 | 5.01** | 1.12 | +1.7 | 1.42 | .77 |
| 12. | I deal with stress and pressure better. | +1.3 | 1.23 | -0.4 | 1.19 | 4.93** | 1.40 | +1.2 | 1.20 | .72 |
| Scale Score | | +15.2 | 11.8 | -4.3 | 12.1 | 11.53** | 1.64 | +15.4 | 12.6 | |
| Internal Consistency (Cronbach's alpha) | | | .94 | | .91 | | | | .96 | |

**$p < .01$; $p < .05$

**Table 3:**
Intercorrelations between the Direct Measure and Indirect Measures of Change in Psychotherapy Patients (Studies 1 and 2)

| Indirect Measure of Change (Difference Scores) | Direct Measure of Change (Q-PC) |
|---|---|
| *75 Psychotherapy Outpatients (Treatment: 12 Weeks; Study 1)* | |
| A-SYM-1: Physical and Mental Stress Symptoms | .45** |
| A-SYM-2: Nervousness and Inner Strain | .38** |
| A-SYM-3: Psychophysiological Deregulation | .32** |
| A-SYM-4: Behavior and Achievement Disorders | .41** |
| A-SYM-5: Pain Distress | .27* |
| A-SYM-6: Problems in Self-Control | .43** |
| A-SYM-G: General Complaints and Symptoms (sum of 1-6) | .48** |
| BDI: Depression | .42** |
| H-Scale: Hopelessness | .49** |
| DSM-IV Axis 5: Global Assessment of Functioning (GAF) | .52** |
| *138 Psychotherapy Outpatients (Treatment: 14-20 Weeks; Sub-Group in Study 2)* | |
| A-SYM-1: Physical and Mental Stress Symptoms | .40** |
| A-SYM-2: Nervousness and Inner Strain | .32** |
| A-SYM-3: Psychophysiological Deregulation | .36** |
| A-SYM-4: Behavior and Achievement Disorders | .44** |
| A-SYM-5: Pain Distress | .19* |
| A-SYM-6: Problems in Self-Control | .37** |
| A-SYM-G: General Complaints and Symptoms (sum of 1-6) | .46** |
| NEO-FFI: Neuroticism | .35** |
| NEO-FFI: Extraversion | .16 |
| NEO-FFI: Openness to Experience | .21* |
| NEO-FFI: Agreeablessness | .26** |
| NEO-FFI: Conscientiousness | .05 |
| ZVT: Cognitive Speed | .19* |
| WAIS - Digit Span: Short Term Memory | .23** |

**p < .01; * $p < .05$

**Table 4:**
Intercorrelations of the Direct Measure of Change (Q-PC) with Response Sets in Study 1 and Study 2

| Response Set | Direct Measure of Change (Q-PC) |
|---|---|
| *Social Desirability (MCSD)* | |
| Study 1: 75 Psychotherapy Outpatients (Treatment: 12 weeks) | .24* |
| Study 2: 138 Psychotherapy Outpatients (Treatment: 14-20 weeks) | .31** |
| *Lying-Scale (EPI-L)* | |
| Study 1: 75 Psychotherapy Outpatients (Treatment: 12 weeks) | .18 |
| Study 2: 138 Psychotherapy Outpatients (Treatment: 14-20 weeks) | .11 |
| *Openness-Scale (FPI-R)* | |
| Study 1: 75 Psychotherapy Outpatients (Treatment: 12 weeks) | -.08 |
| Study 2: 138 Psychotherapy Outpatients (Treatment: 14-20 weeks) | .05 |

**$p < .01$; *$p < .05$

Since the short "Questionnaire of Personal Changes" (Q-PC) is not factor-analytically constructed, there are no hypotheses on the factor structure of its items. Hence, a posteriori standard exploratory factor analysis with varimax rotation was computed. The resulting three factors ($e_{ij} > 1.00$) explain 58 % of the total variance. Factor I (relative variance: 83 %) is marked by considerable factor loadings ($a_{ij} \geq .55$) of nine Q-PC items, Factor II (11 %) is marked by item 6 ($a_{ij} = .44$) and item 12 ($a_{ij} = .39$), and Factor III (6 %) by item 9 ($a_{ij} = .65$).

In clinical psychology we are in need of empirically proven statistical indicators for determining significant and meaningful clinical changes. Intraindividual changes are evaluated accurately by critical change Q-PC scores, which must be exceeded (or fallen below) to be evaluated as a statistically significant improvement (or deterioration). It is assumed that clinical changes (which are determined by the treatment) are only meaningful when they exceed significant changes observed in the randomized waiting-list control group without treatment. Therefore, the critical score limits and critical score intervals were computed with reference to the standard deviation and reliability coefficient as well as the resulting standard error of measurement and confidence interval in the randomized waiting-list control group for different levels of significance. These intraindividual change norms of the Q-PC score are summarized in the left columns of Table 5.

In addition, the absolute and relative frequency of changes observed in the psychotherapy group and the waiting-list control group, which are observed with reference to the statistical criterion (intraindividual norms) of critical, significant score limits for single case change evaluations, are presented in the four right columns of Table 5. Significant improvements are observed in 80 % of the psychotherapy patients and in 13 % of the

**Table 5:**
Absolute and Relative Frequencies of Change in the Direct Measure of Change (Intraindividual Change Norms) in 75 Psychotherapy Patients and 75 Waiting-List Controls (Study 1)

| Criteria for Change | Critical Interval of Direct Measure of Change | Therapy Patients | | Waiting Controls | |
|---|---|---|---|---|---|
| | | *f* | % | *f* | % |
| Improvement at $p < .001$ | + 17 to + 36 | 39 | 52 % | 8 | 11 % |
| Improvement at $p < .01$ | + 9 to +16 | 13 | 17 % | 0 | 0 % |
| Improvement at $p < .05$ | + 7 and +8 | 6 | 8 % | 2 | 3 % |
| Improvement at $p < .10$ | +6 | 2 | 3 % | 0 | 0 % |
| No Significant Change ($p > .10$) | +5 to -5 | 15 | 20 % | 27 | 36 % |
| Deterioration at $p < .10$ | -6 | 0 | 0 % | 11 | 15 % |
| Deterioration at $p < .05$ | -7 and -8 | 0 | 0 % | 5 | 7 % |
| Deterioration at $p < .01$ | -9 to -11 | 0 | 0 % | 3 | 4 % |
| Deterioration at $p < .001$ | -12 to -36 | 0 | 0 % | 19 | 25 % |
| Chi-Square ($df = 8$) | | | 78.88** | | |
| Criteria for Change (Pooled for + / 0 / -) | Critical Interval of Direct Measure of Change | Therapy Patients | | Waiting Controls | |
| | | *f* | % | *f* | % |
| Significant Improvement | + 6 to + 36 | 60 | 80 % | 10 | 13 % |
| No Significant Change | + 5 to -5 | 15 | 20 % | 27 | 36 % |
| Significant Deterioration | -6 to -36 | 0 | 0 % | 38 | 51 % |
| Chi Square ($df = 2$) | | | 77.14** | | |

**$p < .01$

waiting-list controls--the latter of which can be attributed to the commonly observed quota of spontaneous remission in outpatients with mental disorders. The improvement is statistically significant at $p < .001$ in 52 %, at $p < .01$ in 17 %, at $p < .05$ in 8 %, and at $p < .10$ in 3 % of the psychotherapy group. Statistically significant deteriorations are not observed in the therapy group, but in 51 % of the waiting-list controls. After three months of treatment, a significant change was not found in 20 % of the psychotherapy patients, and after three months of waiting in 36 % of the waiting-list controls. Chi-square tests confirm the significance of these differences in the frequency distributions of the Q-PC change scores between the psychotherapy group and the waiting-list control group (see Table 5).

### Discussion of Study 1

The empirically rigorous construction of a brief direct measure of psychotherapeutic outcomes shows encouraging results in the experimental study with randomization. However, the psychometric characteristics of the brief scale must be cross-analyzed in larger samples. This should include cross-analyses of the scale's reliability and its sensitivity for the measurement of changes as well as validation analyses referring not only to its convergent validity, but also to the scale's discriminative validity.

## Study 2

Psychometric quality of the Q-PC is empirically cross-analyzed in Study 2 to determine its statistical item parameters, scale score reliability, item and scale score sensitivity for changes in response to psychotherapy as well as to selected aspects of its convergent and discriminative validity including response sets. Study 2 follows a naturalistic design that includes a larger sample of psychotherapy outpatients. Convergent validation refers to hypotheses on significant correlations of the direct measure of change to (indirect) pre-test-posttest-differences measures of general complaints and symptoms (replication analysis of results of Study 1) as well as neuroticism. Furthermore, the hypothesis is tested that these convergent validity correlation coefficients are significantly stronger than the discriminative validity correlations to extraversion, openness to experience, agreeablessness, conscientiousness, cognitive speed, and short-term memory, that is, variables which are not core outcome concepts of psychotherapy.

### Method

*Participants.* Participants were an unselected sample of 275 adult outpatients of experienced psychotherapists in private practice. Primary diagnoses show a broad spectrum of mental disorders representing the most frequent disorders in outpatient psychotherapy in the geographical region included in the study. Again, depression disorders (DSM-IV: 296.xx, $n$ = 101), anxiety disorders (DSM-IV: 300.2x and 300.02, $n$ = 79), adjustment and stress disorders (DSM-IV: 309.xx, $n$ = 44), and somatoform disorders (DSM-IV: 308.x, $n$ = 24) were most frequently observed. In addition, there were some patients with eating disorders (DSM-IV: 307.1 and 307.51, $n$ = 12), somatoform pain disorders (DSM-IV: 307.8x, $n$ = 10), and hypochondriasis (DSM-IV: 300.7, $n$ = 5). Age of patients ranged from 19 to 69 years ($M$ = 36.7, $SD$ = 13.20). There are 164 females and 111 males in the sample. Physical examinations excluded medical factors. Occupational status and level of education indicate that the majority of the patients were middle class.

After being informed about the treatment process in personal interviews, patients provided their individual consent to psychotherapy and to the diagnostic and evaluation procedures by signing informed consent forms with their full names. Moreover, patients were guaranteed that their records (for study purposes) would be recorded anonymously

and that later rescinding of consent will have no effect on the treatment. Ten patients did not give their consent to participate in the study and are not included in the sample.

*Psychotherapists.* Psychotherapies were conducted by 18 experienced psychotherapists in private practice (job experience: 11-32 years) and were covered by the patient's health insurance. All had professional licensing and full certifications in cognitive-behavioral psychotherapy ($n$ = 16), psychodynamic therapy ($n$ = 4), client-centered psychotherapy ($n$ = 7), and/or relaxation therapy ($n$ = 17). Their basic therapeutic orientation is based on the general psychological therapy approach (Grawe, 2004) focusing on individualized and adaptive treatment, which was professionally supervised on a regular basis (at least once every two weeks).

*Procedure.* A naturalistic design was employed with posttests administered to all patients 12 to 20 weeks after treatment begin (with approximately one session weekly). Pretests were administered only in subgroups (see below).

*Measures.* Initial stage clinical interviews according to DSM-IV were conducted by the therapists with SCID-I and SCID-II (First et al., 1996). In addition, for convergent and discriminative validity analyses, pre- and posttests were conducted with the *German Symptom Checklist* (A-SYM; Krampen, 2007) measuring physical and mental stress (see Study 1; split-half reliability in the present sample: $r_{tt} > .79$), the German Version of the *NEO Five Factor Inventory* (NEO-FFI; Costa & McCrae, 1992; Borkenau & Ostendorf, 1993; $r_{tt} > .75$), the German version of the subtest Digit Span from the *Wechsler Adult Intelligence Scale* (WAIS-R; Wechsler, 1981; Tewes, 1991; $r_{tt} > .69$), and a German *test of cognitive speed* (Zahlen-Verbindungs-Test, ZVT; Oswald & Roth, 1987; $r_{tt} > .77$) in 138 of the patients. The dropout rate during treatment was 9 %, that is, the data of 13 patients are not included in further analyses.

As in Study 1, posttest data on response sets like social desirability *(Marlowe-Crowne Desirability Scale, MCDS;* Crowne & Marlowe, 1964; $r_{tt}$ = .79), faking (L-Scale of the *Eysenck Personality Inventory, EPI;* Eysenck, 1970; $r_{tt}$ = .75), and openness in questionnaire responses (O-Scale of the *Freiburger Persönlichkeitsinventar, FPI-R;* Fahrenberg et al., 2001; $r_{tt}$ = .80) were gathered in the sample of 138 psychotherapy outpatients.

In addition, all participants responded to the Questionnaire of Personal Changes (Q-PC; see Appendix) 14 to 20 weeks after beginning their psychotherapy (13-22 sessions).

## Results

Means and standard deviations of the brief scale's items on personal changes are summarized for the total sample of Study 2 in the right columns of Table 2. For all 12 items there are positive ratings of change in the psychotherapy patients. These results confirm--in a naturalistic design--the sensitivity of the items to specific changes following psychotherapy. The items sum up in a scale score sensitive to specific changes after psychotherapy: Mean scale score is $M$ = +15.4, which is similar to the mean observed in the psychotherapy group in Study 1. The same is true for the internal consistency of the Q-PC score; item-total correlations of all items are acceptable as well (see Table 2).

Intercorrelations between Q-PC scale score and indirect measures of change (difference scores) are summarized in the lower part of Table 3. The results on Q-PC score's correlations to pretest-posttest differences in the A-SYM scales on subjective complaints and symptoms replicate the results of Study 1 in an independent sample. As hypothesized, pretest-posttest differences in neuroticism are significantly correlated to the direct measure of change scale as well, thus confirming convergent validity. Discriminative validity of the Q-PC scale is confirmed by rather low, in part, nonsignificant correlations to pretest-posttest differences in the two indicators of cognitive functioning (cognitive speed and short-term memory) as well as the personality factors extraversion, conscientiousness, openness to experience, and agreeablessness. The difference between the two convergent validity correlations (Q-PC with A-SYM-G and neuroticism) and the six discriminative validity correlations (Q-PC with extraversion, openness to experience, agreeablessness, conscientiousness, cognitive speed, and short-term memory) is statistically significant (using Fisher's z-transformation, $/d/ = .24$, $p < .05$).

The correlations of the Q-PC scale score to the response set scales' scores are--in good convergence with the results of Study 1--rather low. Faking (Lying Scale) and openness in questionnaire responding (O-Scale) are not significantly correlated, the correlation of social desirability (MCDS) with the direct measure of personal changes in psychotherapy outpatients is statistically significant yet low (common variance: 9.6 %; see Table 4). Thus, all results of Study 1 are replicated in the independent sample of Study 2.

As in Study 1, a posteriori standard exploratory factor analysis with varimax rotation was computed. This resulted in two factors ($e_{ij} > 1.00$) explaining 54 % of the total variance. Factor I (relative variance: 88 %) is marked by considerable factor loadings ($a_{ij} \geq .56$) of 10 Q-PC-items, Factor II (12 %) is composed of item 2 ($a_{ij} = .69$) and item 7 ($a_{ij} = .47$).

Regarding the intraindividual norms for the significance of changes (which were computed in Study 1 by the data of the waiting-list control group), for 75 % of the 275 patients having participated in Study 2, significant improvements in the direct measure of change (Q-PC) were observed, for 25 % of them there was no significant change observed, and finally, a deterioration was observed for none of the patients. Specifically, 50 % indicated subjective improvement in contrast to waiting-list controls at a significance level of $p < .001$, 15 % at $p < .01$, 6 % at $p < .05$, and 4 % at $p < .10$.

## General discussion

The results presented here on the experimental construction and cross-validation of the "Questionnaire of Personal Changes" (Q-PC) are promising. With reference to an intervention-based scale construction, the Q-PC's item and scale score sensitivity to change in response to psychotherapy in adult outpatients were confirmed by applying an experimental design with randomization (RCTs) of patients to psychotherapy or a waiting-list control group. Items' and scale score's sensitivity for change were computed by comparisons of specific changes in the treatment group following psychological intervention and the waiting-list control group without intervention. Instrumental reliability and good

psychometric item parameters are empirically demonstrated, and these findings were replicated in the independent Study 2 following a naturalistic design.

Also promising are the results of convergent and discriminative validation found in both studies. There are crucial, significant correlations of the direct change measure to pretest-posttest differences in scales of symptoms, depression, hopelessness, and neuroticism with these being indicators of significant indirect measures of change during and after psychotherapy. Furthermore, the convergent validity of the Q-PC is not only confirmed by the repeated measurement of the variables in patients, but with therapists' repeated global assessments of patients' functioning (GAF) as well, a finding which indicates the value of multimodal assessment of psychotherapy outcomes (see, e.g., Monson, Gradus, Young, Schnurr, Price, & Schumm, 2008). Convergent validation coefficients range between 10 % and 27 % thus revealing the complementary value of the direct measure of change in psychotherapy patients to indirect measures of change: Both strategies of change measurement are not alternative strategies, rather they are complementary in the approximation of the ″true″ changes patients with mental disorders experience in psychotherapy.

Discriminative validity of the Q-PC scale was confirmed for four of the Big Five personality factors and for two indicators of cognitive skills that are hypothesized to remain uninfluenced by psychotherapy. The difference between convergent and discriminative validity correlations is statistically significant. In addition, for response strategies like faking and openness in responding to questionnaire items, no significant correlations to the Q-PC scores in the two independent samples of psychotherapy outpatients were obtained. However, social desirability displays a rather consistent, yet low correlation with the direct measure of change in psychotherapy patients.

A posteriori computed standard explorative factor analyses with varimax rotation led to inconsistent results. However, it must be noted that the brief "Questionnaire of Personal Changes" (Q-PC) is not factor-analytically constructed and that there are no hypotheses on the factor structure of its items. Its factor structure varies with the sample. The results suggest a rather homogeneous factor structure with a strong first factor and one or two more factors with only one or two considerable factor loadings.

Future research should analyze the components of subjective judgments of change in patients and therapists alike. Judgments are comparative in time and may be settled on contrasts of different subjectively experienced discrepancies: They can refer to (1) direct experiences of change without any search in the autobiographic memory and without attributions by others (i.e., criterion-oriented comparisons with reference to quality marks), (2) intraindividual temporal comparisons of the present state to mental representations of some time prior to starting psychotherapy (i.e., autobiographical memory), and (3) direct or indirect attributions of change by others (i.e., attributions of therapist, family members, friends, colleagues, etc.). Thus far, change-sensitive items demonstrate subjectively experienced changes in response to psychotherapy, but it remains unclear about exactly what the foundations of these judgments are.

Furthermore, the question of which relative changes across different treatment groups do occur should be examined in detail (the present data does not provide this information).

This could be, for instance, combined with a survival analysis approach. A good application of this approach has been demonstrated by Luke and Homan (1998) with an illustrative data set from a fictitious study of alcohol relapse patterns comparing survival probabilities for different groups of patients. This is in line with some of the arguments presented by Nesselroade and Ghisletta (2003) for the structuring and measuring of change over the life span in developmental psychology. Besides the methodological problems of change measurement and related modeling strategies, Nesselroade and Ghisletta discuss questions of design as crucial; these refer, for example, to the proper selection of occasions of measurement and to the use of mini-longitudinal studies to overcome practical difficulties in research. In addition, this is in accordance with the use of multiwave data approaches and individual growth modeling in indirect measurement of change (see, e.g., Willett, 1989) for increasing the reliability of change measurements. However, the results presented here on the experimental construction and cross-validation of the "Questionnaire of Personal Changes" (Q-PC) confirm the reliability of a direct measure of change as well. The results are promising for the complementary measurement in the approximation of the ″true″ changes which patients with mental disorders experience in the course of psychotherapy.

## References

Ard, N., & Cook, S. W. (1977). A short scale for the measurement of change in verbal racial attitude. *Educational and Psychological Measurement, 37,* 741-744.

Balzer, D. M. (1989). The direct and indirect measurement of life change. *Dissertation Abstracts International, 49 (12-B),* 5538-5539.

Beck, A. T., Hautzinger, M., Bailer, M., Worall, H., & Keller, F. (1995). *Beck-Depressions-Inventar* [Beck Depression Inventory, BDI]. Göttingen, Germany: Hogrefe.

Beck, A. T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology, 42,* 861-865.

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 3-20). Madison, WI: University of Wisconsin Press.

Beutler, L. E., & Hamblin, D. L. (1986). Individualized outcome measures of internal change: Methodological considerations. *Journal of Counseling and Clinical Psychology, 54,* 48-53.

Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa & McCrae* [NEO Five Factor Inventory after Costa & McCrae]. Göttingen, Germany: Hogrefe.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and NEO Five Factor Inventory.* Odessa, FL: Psychological Assessment Resources.

Cribbie, R., A., & Jamieson, J. (2000). Structural equation models and the regression bias for measuring correlates of change. *Educational and Psychological Measurement, 60,* 893-907.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive.* Ney York, NY: Wiley.

Embretson, S. E. (2006). The continued search for nonarbitrary metrics in psychology. *American Psychologist, 61,* 50-55.

Eysenck, H. J. (1970). *Eysenck Personality Inventory (EPI).* London: University of London Press.

Fahrenberg, J., Hampel, R., & Selg, H. (2001). *Freiburger Persönlichkeitsinventar* [Freiburg Personality Inventory]. Göttingen, Germany: Hogrefe.

First, M. B., Gibbon, M., Spitzer, R. L., & Williams, J. B. E. (1996). *User's Guide for the Structured Clinical Interview for DSM-IV Axis 1 Disorders – Research Version (SCID-I).* Washington, DC: American Psychiatric Press.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. E. (1996). *User's Guide for the Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II).* Washington, DC: American Psychiatric Press.

Foerster, F. (1995). On the problem of initial-value dependencies and measurement of change. *Journal of Psychophysiology, 9,* 324-341.

Grawe, K. (2004). *Psychological therapy.* Cambridge, MA: Hogrefe & Huber.

Hsu, L. M. (1995). Regression toward the mean associated with measurement error and the identification of improvement and deterioration in psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 141-144.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12-19.

Jamieson, J. (1995). Measurement of change and the law of initial values: A computer simulation study. *Educational and Psychological Measurement, 55,* 38-46.

Johnson, E. K., Dow, C., Lynch, R. T., & Hermann, B. P. (2006). Measuring clinical significance in rehabilitation research. *Rehabilitation Counseling Bulletin, 50,* 35-45.

Krampen, G. (1994). *Skalen zur Erfassung von Hoffnungslosigkeit* [Scales for the measurement of hopelessness]. Göttingen, Germany: Hogrefe.

Krampen, G. (2007). *Änderungssensitive Symptomliste zu Entspannungserleben, Wohlbefinden, Beschwerden- und Problembelastungen* [Change sensitive symptom checklist on relaxation, wellness, complaints, and problems]. Göttingen, Germany: Hogrefe.

Lambert, M. J., & Hill, C. E. (1994). Assessing psychotherapy outcomes and processes. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 72-113). Oxford, England: Wiley.

Luke, D. A., & Homan, S. M. (1998). Time and change: Using survival analysis in clinical assessment and treatment evaluation. *Psychological Assessment, 10,* 360-378.

May, K., & Hittner, J. B. (2003). On the relation between power and reliability of difference scores. *Perceptual and Motor Skills, 97,* 905-908.

Miller, E. (1992). Some basic principles of neuropsychological assessment. In J. R. Crawford, D. M. Parker, & W. W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 7-20). Hillsdale, NJ: Lawrence Erlbaum.

Minz, J., Luborsky, L., & Christoph, P. (1979). Measuring outcome of psychotherapy. *Journal of Consulting and Clinical Psychology, 47,* 319-334.

Monson, C. M., Gradus, J. L., Young-Xu, Y., Schnurr, P. P., Price, J. L., & Schumm, J. A. (2008). Change in posttraumatic stress disorder symptoms: Do clinicians and patients agree? *Psychological Assessment, 20,* 131-138.

Nesselroade, J. R., & Ghisletta, P. (2003). Structuring and measuring change over the life span. In U. Lindenberger & U. M. Staudinger (Eds.), *Understanding human development* (pp. 317-337). Dordrecht, Netherlands: Kluwer.

Oswald, W. D., & Roth, W. (1987). *Der Zahlen-Verbindungs-Test (ZVT)* [Digits-Connection-Test]. Göttingen, Germany: Hogrefe.

Reise, S. P., & Haviland, M. G. (2005). Item response theory and the measurement of change. *Journal of Personality Assessment, 84,* 228-238.

Sandell, R. (1987). Assessing the effects of psychotherapy. *Psychotherapy and Psychosomatics, 47,* 29-36.

Shaffer, P., Murillo, N., & Michael, W. B. (1980). The factor dimensions of a revised scale for student client evaluation of counselors in a university counseling and testing center. *Educational and Psychological Measurement, 40,* 1081-1089.

Tan, E. S., Imbos, T., Does, R. J. M. M., & Theunissen, M. (1995). An optimal, unbiased classification rule for mastery testing on longitudinal data. *Educational and Psychological Measurement, 55,* 595-612.

Tewes, M. (1991). *Hamburg-Wechsler-Intelligenztest für Erwachsene – Revision* [Hamburg-Wechsler Intelligence Test for Adults – Revision]. Göttingen, Germany: Hogrefe.

Vermeersch, D. A. (1998). Sensitivity to change as an item selection criterion for an outcome measure. *Dissertation Abstracts International, 59 (6-B),* 3117.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale – Revised.* New York, NY: Psychological Corporation.

Weinstock, M. C., & Meier, S. T. (2003). A comparison of two item-selection methodologies for measuring of change in university counseling center clients. *Measurement and Evaluation in Counseling and Development, 36,* 66-75.

Zielke, M. (1980). Darstellung und Vergleich von Verfahren zur individuellen Veränderungsmessung [Description and comparison of methods assessing individual change]. *Psychologische Beiträge, 22,* 592-609.

# Appendix

## Questionnaire of personal changes (Q-PC)

Your name /Code: _____ Date: _____

Please think back to the time prior to beginning your treatment.

**The treatment started on** _____ (please enter the date!),
**approximately _____ weeks ago** (please enter the number of weeks!).

Try to remember what you did at that time and how you felt.

For the following statements, please indicate the kind of changes you experienced, in one direction or the other, during your treatment. Use the scale provided below to evaluate these changes.

+3 = strong positive change
+2 = medium positive change
+1 = weak positive change
0 = no change
–1 = weak negative change
–2 = medium negative change
–3 = strong negative change

| | +3 | +2 | +1 | 0 | –1 | –2 | –3 |
|---|---|---|---|---|---|---|---|
| 01. I can relax much better. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| 02. I can unwind better and take it easy. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| 03. Overall I feel healthier. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| 04. I feel less anxious thinking about the future. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| 05. I feel calmer and more well-balanced. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| 06. I sleep better. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| 07. I take less medication. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| 08. I have more stamina and do not give up as easily. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| 09. I can concentrate much better. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| 10. I cope with unexpected events more easily. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| 11. I feel better. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| 12. I deal with stress and pressure better. | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |