

Psychometric Properties of the German Version of the Scale to Assess the Therapeutic Relationship in Community Mental Health Care (D-STAR)

Sabine Loos¹, Reinhold Kilian¹, Thomas Becker¹, Birgit Janssen², Harald Freyberger³, Hermann Spiessl⁴, Julia Grempler⁵, Stefan Priebe⁶, and Bernd Puschner¹

¹Department of Psychiatry II, Ulm University, Günzburg, Germany, ²Department of Psychiatry and Psychotherapy, Düsseldorf University, Germany, ³Department of Psychiatry and Psychotherapy, Greifswald University, Stralsund, Germany, ⁴Department of Psychiatry and Psychotherapy, Regensburg University, Germany, ⁵Department of Psychiatry and Psychotherapy I, Ulm University, Ravensburg, Germany, ⁶Unit for Social and Community Psychiatry, Queen Mary University of London, UK

Abstract. *Objective:* There are presently no instruments available in German language to assess the therapeutic relationship in psychiatric care. This study validates the German version of the Scale to Assess the Therapeutic Relationship in Community Mental Health Care (D-STAR). *Method:* 460 persons with severe mental illness and 154 clinicians who had participated in a multicenter RCT testing a discharge planning intervention completed the D-STAR. Psychometric properties were established via item analysis, analyses of missing values, internal consistency, and confirmatory factor analysis. Furthermore, convergent validity was scrutinized via calculating correlations of the D-STAR scales with two measures of treatment satisfaction. *Results:* As in the original English version, fit indices of a 3-factor model of the therapeutic relationship were only moderate. However, the feasibility and internal consistency of the D-STAR was good, and correlations with other measures suggested reasonable convergent validity. *Conclusions:* The psychometric properties of the D-STAR are acceptable. Its use can be recommended in German-speaking countries to assess the therapeutic relationship in both routine care and research.

Keywords: psychometrics, therapeutic relationship, STAR, community mental health care

A beneficial patient-therapist relationship has been recognized as an important process factor in psychotherapy research, regardless of the specific treatment method or setting variables (Elvins & Green, 2008; Martin, Garske, & Davis, 2000; Nocross, 2002). Over the last few decades, the therapeutic relationship (TR) has been studied extensively in psychotherapy via a number of instruments based upon various theoretical backgrounds (Catty, 2004; McCabe & Priebe, 2004). The last years have seen a continuous diversification of the concept of the TR including the application outside its “genuine territory,” which is individual outpatient psychotherapy. Specifically, researchers have begun to acknowledge the importance of the TR in

inpatient and community psychiatric treatment (Howgego, Yellowlees, Owen, Meldrum, & Dark, 2003; McGuire, McCabe, & Priebe, 2001; Priebe & McCabe, 2006). However, TR research approaches originating in psychotherapy cannot be easily transferred to the field of psychiatry, which differs from psychotherapy in many aspects. Compared to psychotherapy clients, psychiatric patients generally show higher levels of severity and chronicity of illness, and are usually treated by multiprofessional teams with a primary focus of treatment on attaining stability rather than “genuine” psychotherapeutic treatment goals such as changing cognitive and emotional processes (Priebe & Gruyters, 1993; Priebe & McCabe, 2006).

Thus, adaptations are required regarding models and instruments to measure the TR in psychiatric care (Catty, Winfield, & Clement, 2007; Howgego et al., 2003). However, there is a lack of measures for assessing the TR in psychiatric treatment settings. Instruments originating in psychotherapy research are commonly used, with the Working Alliance Inventory (WAI; Horvath & Greenberg, 1989) being especially prominent (Bale, Catty, Watt, Greenwood, & Burns, 2006; McCabe, Saidi, & Priebe, 2007). With some exceptions (Blais, 2004; Clarkin, Hurt, & Crilly, 1987; Priebe & Gruyters, 1993; Pulido, Monaria, & Rossia, 2008; Stark, Lewandowski, & Buchkremer, 1992), attempts to develop measures of the TR for explicit use in psychiatry have been rare, and such genuine psychiatric TR measures were developed on an ad-hoc basis, are extremely reductionist, and lack appropriate establishment of psychometric properties (McCabe, Roder-Wanner, Hoffmann, & Priebe, 1999). Until recently, no standardized measure was available to comprehensively assess the quality of the TR in psychiatric community care. To fill this gap, McGuire-Snieckus, McCabe, Catty, Hansson, and Priebe (2007) introduced the Scale to Assess the Therapeutic Relationship (STAR) based upon Bordin's (1979) pantheoretical model of alliance.

In this paper we introduce the German version of this scale (D-STAR-G) by making the translation procedure transparent and testing its psychometric properties in a sample of $N = 460$ people with severe mental illness who had been discharged from psychiatric inpatient care. Specific research questions were (1) to examine the feasibility of the D-STAR in terms of item analysis and reliability, (2) to test whether the 3-factor model for the D-STAR-C and the D-STAR-P suggested by the authors of the original version could be confirmed in a German patient sample, and (3) to scrutinize the D-STAR's convergent validity.

Material and Methods

Design

Between April 2006 and July 2007, 953 users of mental health inpatient services at five psychiatric hospitals in Germany (Günzburg, Düsseldorf, Greifswald, Regensburg, and Ravensburg) were invited to participate in the study entitled "Effectiveness and Cost-Effectiveness of Needs-Oriented Discharge Planning and Monitoring for High Utilizers of Psychiatric Services" (NODPAM). NODPAM was a multi-center randomized controlled trial with four measurement points: baseline (T0, at hospital discharge), and 3 (T1), 6 (T2), and 18 months (T3) thereafter. Inclusion criteria were (1) age 18–65 years, (2) a primary diagnosis of schizophrenia, bipolar affective disorder, or major depression as recorded by the clinician at admission according to ICD-10 criteria, and (3) previous high utilization of psychiatric inpatient care, defined as occurring during the 2 years prior to current inpatient treatment, (a) at least two admissions with a cumulative length of stay (LOS) exceeding 30 days or (b) at least one admission

with a cumulative LOS of more than 50 days. 491 participants were included after obtaining their informed consent and were randomized to either the intervention group (IG; $N = 241$) or the control group (CG; $N = 250$). 421 of the 491 patients initially included also participated at T1 (209 IG, 212 CG). For a detailed description of the design and method, see Puschner, Steffen et al. (2008).

Instruments

McGuire-Snieckus and colleagues' (2007) clinician (STAR-C) and patient (STAR-P) versions have 12 items and three subscales each. Two of these are identical in both versions: Positive Collaboration (PC) and Positive Clinician Input (PCI). The third subscales are Emotional Difficulties (ED) in the STAR-C, and Nonsupportive Clinician Input (NSCI) in the STAR-P. Translation of the STAR from English to German took place during the preparation of the NODPAM study and pragmatically followed the ten principles of good practice for translations and cultural adaptation of outcome measures (Wild et al., 2005). Forward translation was carried out by a native-speaking English professional translator. Then, a bilingual expert panel (SP, SL, BP) screened the forward translation for consistency and linguistic appraisal, i.e., identified and resolved inadequate expressions or concepts. Next, the revised German version was backtranslated into English by a professional translator and sent to the test authors (RM, SP) for final comparisons with the original. Final minor corrections following notes from the test authors were made before completion of the final German version. Finally, patients and clinicians participating in the NODPAM study were invited to complete the D-STAR at all measurement points (T0–T1). Patients filled in the questionnaire in the presence of a study worker. German wording of the D-STAR items can be obtained from the corresponding author or via the internet (http://www.uni-ulm.de/psychiatrieII/projekte/nodpam_mat.htm).

Treatment satisfaction from patient and clinician perspectives was assessed at T1. Instruments used were the ZUF-8 (Schmidt, Lamprecht, & Wittmann, 1989), which is the German version of the Client Satisfaction Questionnaire (CSQ; Attkisson & Zwick, 1982) for patients, and the Questionnaire to Assess Therapist Work Satisfaction (ZUF-THERA; Puschner, Bauer, Kraft, & Kordy, 2005) for key workers.

Data Analysis

Psychometric properties in terms of feasibility and reliability of the D-STAR were assessed by means of item analyses (mean, standard deviation, skewness, kurtosis, item correlations), analysis of missing values, and consistency coefficients (Cronbach's α). A confirmatory factor analysis (CFA) was performed to test whether the 3-factor structure of both versions of the STAR suggested by the authors of the original version could be replicated in the German sam-

Table 1. Item characteristics

Item	Mean (SD)	Skewness	Kurtosis	Missing %
D-STAR-C				
1. I get along well with my patient.	3.00 (0.58)	-0.37	1.13	0.2
2. My patient and I share a good rapport.	2.79 (0.66)	-0.52	0.60	0.4
3. I listen to my patient.	2.91 (0.67)	-0.60	1.75	1.1
4. I feel my patient rejects me as a clinician. ¹	3.15 (0.71)	-0.63	0.73	0.4
5. I believe my patient and I share a good relationship.	2.90 (0.61)	-0.75	1.75	0.2
6. I feel inferior to my patient. ¹	3.55 (0.64)	-1.52	2.77	0.9
7. My patient and I share similar expectations regarding progress in treatment.	2.46 (0.78)	-0.60	-0.10	0.2
8. I feel that I am supportive of my patient.	3.20 (0.58)	-0.32	1.06	0.4
9. It is difficult for me to empathize with or relate to my patient's problems. ¹	2.78 (0.76)	-0.64	0.49	0.0
10. My patient and I are open with one another.	2.94 (0.61)	-0.55	1.35	0.4
11. I am able to take my patient's perspective.	2.77 (0.58)	-0.90	1.89	0.2
12. My patient and I share a trusting relationship.	2.91 (0.62)	-0.54	1.16	0.2
D-STAR-P				
1. My clinician speaks with me about my personal goals and thoughts about treatment.	2.47 (1.12)	-0.33	-0.64	0.2
2. My clinician and I are open with one another.	3.19 (0.98)	-1.13	0.50	0.9
3. My clinician and I share a trusting relationship.	2.94 (1.18)	-0.96	-0.00	1.5
4. I believe my clinician withholds the truth from me. ¹	2.91 (1.28)	-0.92	-0.31	1.5
5. My clinician and I are honest with one another.	3.23 (0.98)	-1.34	1.26	0.9
6. My clinician and I work toward mutually agreed goals.	3.03 (1.15)	-1.18	0.59	1.1
7. My clinician is stern with me when I speak about things that are important to me. ¹	1.90 (1.43)	0.14	-1.23	2.4
8. My clinician and I have established an understanding of the kind of changes that would be good for me.	2.65 (1.12)	-0.77	-0.01	2.2
9. My clinician is impatient with me.	3.06 (1.23)	-1.17	0.24	0.9
10. My clinician seems to like me regardless of what I do or say.	2.41 (1.32)	-0.45	-0.95	4.1
11. We agree on what is important for me to work on.	2.97 (1.06)	-1.02	0.53	1.3
12. I believe my clinician has an understanding of what my experiences have meant to me.	2.81 (1.13)	-0.93	0.21	2.0

Note: ¹Item recoded for analysis.

ple. CFA was carried out using the maximum likelihood estimation analyzing the covariance matrix of the items. In addition to the χ^2 and χ^2/df ratio (cutoff ≤ 3 ; Homburg & Pflesser, 2000), other global fit indices were checked: the comparative fit index (CFI, cutoff $\geq .90$), the Tucker Lewis index (TLI, cutoff $\approx .95$), the root mean square error of approximation (RMSEA, cutoff $< .06$ – $.08$), and the standardized root mean square residual (SRMR, cutoff $\leq .11$; Fan, Thompson, & Wang, 1999; Hu & Bentler, 1999). Spearman's rank correlations were used to examine the convergent validity (TR and satisfaction with treatment). We conducted the CFA with Mplus 4.1 by Muthén and Muthén (1998–2010).

Results

Sample

460 of the total of 491 study participants completed the D-STAR-P at baseline. Participants' mean age was 41.5

($SD = 11.3$) years, about half of them (48%) were women, 18.4% were married, and more than half (54.8%) had no children. Educational level was predominantly low (37.3 low-track Hauptschule) and one-fifth (22.1%) lived in assisted accommodations. Most were retired (49.9%), and only a minority (16.4%) was in paid employment. Diagnoses were schizophrenia (ICD-10 F2, 58.3%) and affective disorder (F3, 41.7%). Mean time since first psychiatric hospital admission was 8.8 ($SD = 8.3$) years, and mean cumulated LOS during 2 years prior to the index admission was 122.8 ($SD = 89.3$) days. All patients had been discharged from the hospital with medication.

154 inpatient clinicians (psychiatrists or psychologists) who completed the D-STAR-C for 453 patients at baseline were 36.2 ($SD = 8.3$) years old on average, and 78 (58.2%) of them were female. Their overall job experience was 7.4 ($SD = 7.8$) years, 5.1 ($SD = 7.0$) years of which in mental health services. 84 (66.6%) of them were psychiatry residents, 22 (16.9%) were consultants, 17 (13.1%) were psychotherapists in training, and 7 (5.4%) were psychotherapists.

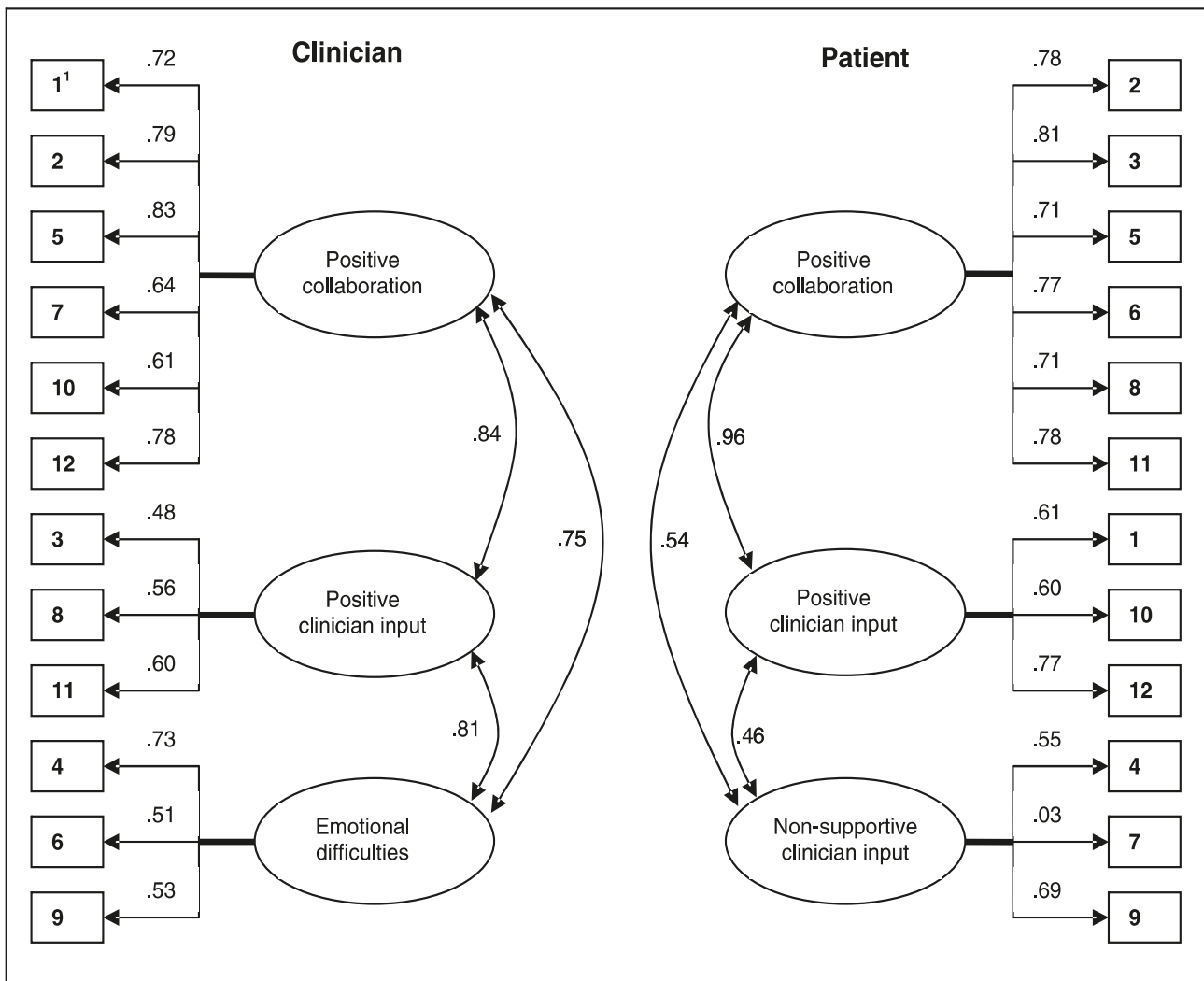


Figure 1. Confirmatory factor solution for the D-STAR. ¹Item number.

Feasibility and Distribution Characteristics

Clinician Version

The mean total score of the D-STAR-C was 35.2 ($SD = 5.2$). The distribution of the item characteristics (Table 1) shows that all items have a homogeneous range ($SD = 0.58-0.78$). Items with the most extreme difficulties were #6 (3.55) and #7 (2.46), which indicates that only about 27% of the whole range of the 5-point Likert scale was used by the majority of participating clinician. The items were left-skewed ($g_1 < 0$) ranging from -0.32 to -1.52 . Average rate of missing values was 0.4% (range 0–1.1%). Furthermore, the lowest response category *never* was not chosen in 7 out of 12 items (#1, #2, #5, #6, #8, #10, and #12). Internal consistency (Cronbach's α) of the scale was 0.87, and the range of discriminative power of all items was between 0.35 and 0.73, indicating an average to high level of item homogeneity.

Patient Version

The mean total score of the D-STAR-P was 32.9 ($SD = 8.8$). Item difficulty varied between 3.23 and 1.90 (Table 1) indicating that only 33% of the entire range of the 5-point Likert scale was used by the majority of participating patients, mostly on the part of agreement. Items were left-skewed ($g_1 < 0$) except for #7. Mean rate of missing values over all items was a bit higher than in the clinician version (1.6%; range 0.2–4.1%). The reliability of the entire scale was moderate ($\alpha = 0.83$), due mainly to the negative discriminative power (-0.28) of item #7. Cronbach's α without this item was 0.88. Discriminatory power of all other items ranged between 0.34 and 0.73.

Confirmatory Factor Analysis

Results of the CFA for the 3-factor solution of the D-STAR clinician/patient versions were as follows: $\chi^2 (df) = 231.81$

(51)/226.47 (51), $p \leq .001 / < .001$; CFI = 0.91/0.92; TLI = 0.86/0.90; RMSEA = 0.09/0.09; and SRMR = 0.05/0.08 (Figure 1). For the D-STAR-C, the χ^2 test was significant, indicating a difference between theoretical and observed relations. Also, other indices failed to achieve optimal model fit. The standardized loadings of the 3 factors were moderate to strong (0.48–0.83). The 3-factor solution for the D-STAR-P showed a similar picture but slightly better fit indices. The standardized loadings of the 3 factors were of moderate to strong (0.55–0.81) except for item #7 (0.03). With $R^2 = 0.001$, this item also accounted for the lowest amount of variance of the three items making up the subscale Nonsupportive Clinician Input (NSCI), indicating that, compared to the other two items, it does not represent the same content.

Correlation between total sum scores of D-STAR-C and D-STAR-P was $r = 0.19$, $p < .001$. Correlations between D-STAR-C total and subscales scores ranged between $r = 0.42$ and $r = 0.93$, for the patient version correlations ranged between $r = 0.07$ and $r = 0.93$ with the subscale NSCI correlating low with other subscales and the sum score.

Convergent Validity

To test convergent validity, the correlations of the ZUF-8 (patient treatment satisfaction) and the ZUF-THERA (treatment satisfaction with treatment given) with the respective versions of the D-STAR at T1 were scrutinized. All correlations of the ZUF-8 with the D-STAR-P total and subscale scores were substantial (sum score: $r = 0.64$, $p \leq .01$; PC: $r = 0.66$, $p \leq .01$; PCI: $r = 0.55$, $p \leq .01$; and NSCI: $r = 0.20$, $p \leq .01$), as were those of the ZUF-THERA with the D-STAR-C (sum score: $r = 0.53$, $p \leq .01$; PC: $r = 0.54$, $p \leq .01$; PCI: $r = 0.39$, $p \leq .01$; and ED: $r = 0.53$, $p \leq .01$). This indicates that all scales measure a similar construct.

Discussion

This study examined psychometric properties of the German adaptation of the D-STAR in a sample of 460 people who had received inpatient treatment for severe mental health problems (patient version D-STAR-P) and their clinicians ($N = 453$; clinician version D-STAR-C). In line with specifications of classical test theory, we carried out item analysis and examined reliability and validity. Furthermore, factorial structure proposed by the authors of the original version (McGuire-Snieckus et al., 2007) was confirmatorily tested. Return rates of patients (460 of 491 = 94%) and of clinicians (453 of 491 = 92%) were outstanding, indicating that the D-STAR is highly accepted. This is probably due mainly to its high face validity and brevity.

For both clinician and patient versions, item analysis revealed for almost all items rather skewed distributions toward a positive perception of the TR, i.e., the full range of possible answers of the items has not been utilized. This is not uncom-

mon for an instrument that asks about personal experiences and satisfaction at the end of a treatment episode. High satisfaction ratings of users of mental health care are the rule rather than the exception (Bjørngaard, Torleif, & Friis, 2010; Puschner et al., 2005). However, there were also differences between clinician and patient ratings of the TR. Clinicians rated the TR slightly more positively than patients did. Higher clinician ratings of the TR had previously been reported, but compared to the original validation sample (McGuire-Snieckus et al., 2007) and to results of a meta-analysis of 53 studies in psychotherapy (Tryon, Blackwell, & Hammel, 2007), the differences found in our sample were rather small, albeit significant. Also correlations between the clinician and patient TR ratings was lower ($r = 0.19$, $p < .001$) than mean correlation found in this meta-analysis ($r = .36$; Tryon et al., 2007). One obvious reason for this difference is that, in psychotherapy, as opposed to psychiatric treatment, patient and therapist spend much more time with each other on a one-to-one basis allowing for more harmonization of their views on the TR.

With overall missing rates of 0.4% in the clinician and 1.6% in the patient version, feasibility of the D-STAR is very satisfactory. Also, with Cronbach's α 's of 0.87 for the clinician and of 0.83 for the patient version total score, internal consistency can be considered good. However, the internal consistency of the D-STAR-C would have been slightly better without item #7. A confirmatory test of the 3-factor model yielded acceptable fit indices comparable to those of the original version. A 1-factor solution model tested did not reveal an improvement in terms of better fit. In general, with correlations between subscales ranging from 0.46 and 0.96, and often low items loading on subscales, which is especially true for the subscales with only very few items, conceptual distinction between the factors is moderate at best. Convergent validity was confirmed by the moderate but significant correlations (r 's = 0.20 to 0.66) between D-STAR total and subscale scores with measures of client and therapist satisfaction. Hardly surprising, the lowest correlation was found between patient satisfaction and the D-STAR-P subscale "Nonsupportive Clinician Input," which also includes the problematic item #7. Results converge in demonstrating problems with this item in the D-STAR-P which – as indicated by a missing value of 2.4 – is difficult to understand and does not reflect the same meaning as other items making up the NSCI subscale. We believe that this is due mostly to the fact that it is very difficult to adequately transfer into German the English wording and meaning of this item ("My clinician is stern with me when I speak about things that are important to me" – it was particularly difficult to find a proper translation for the word "stern"). Apparently, even though a lot of effort has gone into this during the translation process, carry-over of this item was not entirely successful. Given the overall good psychometric properties of the scale, we would consider this a minor drawback that does not make a reformulation of this item necessary at this stage. However, a subsequent validation study might devote some attention to finding a better version.

In general, when interpreting differences between the German and UK version of the STAR, it should be borne in mind

that the sample used for this study differed from the UK sample in several aspects: In the German sample, clinicians were either psychiatrists or psychological psychotherapists, many of whom were still in training; in the UK sample, provider professions were more diverse (mainly nurses and social workers). Furthermore, patients in the German sample had a lower duration of illness (8.8 vs. 21 years) with lower diversity of ethnic background. Findings of the German version might be considered rather robust since they were based upon a much larger (clinicians $N = 154$ vs. $N = 7$; patients $N = 460$ vs. $N = 68$) and less selective sample recruited from different parts of Germany representing both urban and rural catchment areas. Despite these differences in sample characteristics and general differences in mental health service provision between Germany and the UK, the psychometric properties found for both versions were equally satisfying. This means that the STAR works in different mental health care systems, and it also indicates that the nature and quality of the TR is very similar in both countries.

Limitations

Some limitations of this work should be taken into consideration. First, test-retest reliability and other aspects of validity were not determined. One reason for this was that clinicians changed between baseline at hospital discharge and 3-month follow-up (hospital-based vs. office-based psychiatrists and psychotherapists), making it impossible to treat the two measurements as true subsequent assessment points. Second, contrary to the original version, which was tested in a community setting, in this study the D-STAR was evaluated in a sample of people receiving inpatient mental health care. This adds to the differences between the two samples mentioned above, but also shows the robustness of the scale. Third, strictly speaking, generalization of findings is limited to high utilizers of inpatient mental health care with considerable severity and duration of illness. However, given that the D-STAR did well in this population, there is ample reason to believe that this would also be the case if presented to people with less severe mental health problems. Yet, further evidence is needed to confirm this assumption.

Conclusions and Outlook

Overall, this study showed in a large sample that people with severe and longstanding mental health problems were able to give meaningful information about the quality of the relationship to their treatment providers via the German adaptation of the D-STAR. Also psychometric properties of the clinician version (D-STAR-P) were highly satisfactory. Furthermore, by being brief (12 items only, time to complete max. 5 minutes) and highly accepted by clinicians and patients (excellent return rates), the D-STAR, like the original version, shows outstanding feasibility (Slade, Thornicroft, & Glover, 1999).

This means that a validated standardized instrument is now available specifically devised to measure the quality of the TR in psychiatric service settings. While the TR has been a stronghold in psychotherapy research for many years (Horvath & Symonds, 1991; Martin et al., 2000; Orlinsky, Rønnestad, & Willutzki, 2004; Puschner, Wolf, & Kraft, 2008), psychiatric research literature has only rather recently started to recognize its outstanding importance to the quality of delivering treatment (Catty, 2004; Howgego et al., 2003; McGuire et al., 2001). This fills an important gap since mental health services researchers in German-speaking countries no longer need to rely on TR instruments developed in psychotherapy settings whose suitability for psychiatric treatment is questionable as discussed before.

Future research should address remaining problems in factorial structure especially pertaining to the D-STAR-P subscale "Nonsupportive Clinician Input" and make further progress on producing high-quality evidence on the alliance-outcome relation in psychiatric treatment.

Acknowledgments

This work is part of the NODPAM study, which was supported by a grant from the German Research Foundation (DFG) in the Special Program "Clinical Studies" (Grant number BE 2502/3-1). We are grateful to the following colleagues for significant contributions to the NODPAM study: Petra Schneider, Rana Kalkan, Constanze Lahmeyer (Ulm/Günzburg); Wolfgang Gaebel, Meike Ramacher, Helga Zimmer, Katrin Fisahn (Düsseldorf); Carsten Spitzer, Britta Skoeries, Katrin v. Rad, Kathleen Völker (Greifswald); Helmfried E. Klein, Katja Sohla, Veronika Steinkohl, Annette Janner, Verena Fischer, Natalie Philipp (Regensburg); Tilman Steinert, Jan Bergk, Andrea Döbler, Ute Scheck (Ulm/Weißenau); Rainer Muehe, Friedericke Rohlmann (Ulm/Biometrics).

References

- Attkisson, C. C., & Zwick, R. (1982). The Client Satisfaction Questionnaire. *Evaluation and Program Planning*, 5, 233–237.
- Bale, R., Catty, J., Watt, H., Greenwood, N., & Burns, T. (2006). Measures of the therapeutic relationship in severe psychotic illness: A comparison of two scales. *International Journal of Social Psychiatry*, 52, 256–266.
- Bjørngaard, J. H., Torleif, R., & Friis, S. (2010). The impact of mental illness on patient satisfaction with the therapeutic relationship: A multilevel analysis. *Social Psychiatry and Psychiatric Epidemiology*, 42, 803–809.
- Blais, M. A. P. (2004). Development of an inpatient treatment alliance scale. *Journal of Nervous and Mental Disease*, 192, 487–493.
- Bordin, E. (1979). The generalizability of the psychoanalytic concept of working alliance. *Psychotherapy: Theory, Research and Practice*, 16, 252–260.
- Catty, J. (2004). "The vehicle of success": Theoretical and empirical

- perspectives on the therapeutic alliance in psychotherapy and psychiatry. *Psychology and Psychotherapy*, 77, 255–272.
- Catty, J., Winfield, H., & Clement, S. (2007). The therapeutic relationship in secondary mental health care: A conceptual review of measures. *Acta Psychiatrica Scandinavica*, 116, 238–252.
- Clarkin, J. F., Hurt, S. W., & Crilly, J. L. (1987). Therapeutic alliance and hospital treatment outcome. *Hospital and Community Psychiatry*, 38, 871–875.
- Elvins, R., & Green, J. (2008). The conceptualization and measurement of therapeutic alliance: An empirical review. *Clinical Psychology Review*, 28, 1167–1187.
- Fan, X., Thompson, B., & Wang, L. (1999). The effects of sample size, estimation methods, and model specification on SEM fit indices. *Structural Equation Modeling*, 6(1), 56–83.
- Homburg, C., & Pflesser, C. (2000). Konfirmatorische Faktorenanalyse [Confirmatory factor analysis]. In A. Herrmann & C. Homburg (Eds.), *Marktforschung. Methoden – Anwendungen – Praxisbeispiele* (pp. 413–439). Wiesbaden: Gabler.
- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, 36, 223–233.
- Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology*, 38, 139–149.
- Howgego, I. M., Yellowlees, P., Owen, C., Meldrum, L., & Dark, F. (2003). The therapeutic alliance: The key to effective patient outcome? A descriptive review of the evidence in community mental health case management. *Australian and New Zealand Journal of Psychiatry*, 37, 169–183.
- 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.
- Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68, 438–450.
- McCabe, R., & Priebe, S. (2004). The therapeutic relationship in the treatment of severe mental illness: A review of methods and findings. *International Journal of Social Psychiatry*, 50, 115–128.
- McCabe, R., Roder-Wanner, U. U., Hoffmann, K., & Priebe, S. (1999). Therapeutic relationships and quality of life: Association of two subjective constructs in schizophrenia patients. *International Journal of Social Psychiatry*, 45, 276–283.
- McCabe, R., Saidi, M., & Priebe, S. (2007). Patient-reported outcomes in schizophrenia. *The British Journal of Psychiatry*, 191(50), s21–s28.
- McGuire, R., McCabe, R., & Priebe, S. (2001). Theoretical frameworks for understanding and investigating the therapeutic relationship in psychiatry. *Social Psychiatry and Psychiatric Epidemiology*, 36, 557–564.
- McGuire-Snieckus, R., McCabe, R., Catty, J., Hansson, L., & Priebe, S. (2007). A new scale to assess the therapeutic relationship in community mental health care: STAR. *Psychological Medicine*, 37(1), 85–95.
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus users's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén. Retrieved from <http://www.statmodel.com/download/usersguide/Mplus%20Users%20Guide%20v6.pdf>
- Nocross, J. C. (2002). *Psychotherapy relationships that work: Therapist contributions and responsiveness to patient needs*. New York: Oxford University Press.
- Orlinsky, D. E., Rønnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: Continuity and change. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 307–389). New York: Wiley.
- Priebe, S., & Gruyters, T. (1993). The role of the helping alliance in psychiatric community care. A prospective study. *Journal of Nervous and Mental Disease*, 181, 552–557.
- Priebe, S., & McCabe, R. (2006). The therapeutic relationship in psychiatric settings. *Acta Psychiatrica Scandinavica Supplement*, 113(s429), 69–72.
- Pulido, R., Monaria, M., & Rossia, N. (2008). Institutional therapeutic alliance and its relationship with outcomes in a psychiatric day hospital program. *Archives of Psychiatric Nursing*, 22, 277–287.
- Puschner, B., Bauer, S., Kraft, S., & Kordy, H. (2005). Zufriedenheit von Patienten und Therapeuten mit ambulanter Psychotherapie [Patient and therapist satisfaction in outpatient psychotherapy]. *Psychotherapie, Psychosomatik, medizinische Psychologie*, 55, 517–526.
- Puschner, B., Steffen, S., Gaebel, W., Freyberger, H., Klein, H. E., Steinert, T., ... Becker, T. (2008). Needs-oriented discharge planning and monitoring for high utilizers of psychiatric services (NODPAM): Design and methods. *BMC Health Services Research*, 8, 152.
- Puschner, B., Wolf, M., & Kraft, S. (2008). Helping alliance and outcome in psychotherapy: What predicts what in routine outpatient treatment? *Psychotherapy Research*, 18, 167–178.
- Schmidt, J., Lamprecht, F., & Wittmann, W. W. (1989). Zufriedenheit mit der stationären Versorgung: Entwicklung eines Fragebogens und erste Validitätsuntersuchungen [Satisfaction with inpatient management: Development of a questionnaire and initial validity studies]. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 39, 248–255.
- Slade, M., Thornicroft, G., & Glover, G. (1999). The feasibility of routine outcome measures in mental health. *Social Psychiatry and Psychiatric Epidemiology*, 34, 243–249.
- Stark, F. M., Lewandowski, G., & Buchkremer, G. (1992). Therapist-patient relationships as a predictor of the course of schizophrenia illness. *European Psychiatry*, 7, 161–169.
- Tryon, G. S., Blackwell, S. C., & Hammel, E. F. (2007). A meta-analytic examination of client-therapist perspectives of the working alliance. *Psychotherapy Research*, 17, 629–642.
- Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health*, 8, 94–104.

Published online: December 15, 2011

Bernd Puschner

Department of Psychiatry II, Ulm University
Ludwig-Heilmeyer-Str. 2
89312 Günzburg
Germany
Tel. +49 8221 962866
Fax +49 8221 9628160
E-mail bernd.puschner@bkh-guenzburg.de