Research Articles // Artículos de investigación

Notes and Editorial Information // Avisos e información editorial

IJP&PT

INTERNATIONAL JOURNAL OF

PSYCHOLOGY & PSYCHOLOGICAL

THERAPY

http://www.ijpsy.com

Multi-informant Assessment of Therapeutic Competence: Development and Initial Validation of a Set of Measurements

Christine Koddebusch*, Christiane Hermann

Justus Liebig University, Deutschland

ABSTRACT

Assessment of therapeutic competence is required to be multi-informant and should therefore include the perspectives of therapists, clients and observers. There are several instruments assessing therapeutic competence from one of the requested perspectives. But heretofore there is no valid and reliable set of measurements that allows the assessment of therapeutic competence from the different perspectives based on the same theoretical basis, what is central for making comparisons possible. We developed a set of measurements assessing therapeutic competences from different perspectives: therapists' global (GloRa-T) and session self-rating (SeRa-T), clients' session rating (SeRa-C) and observer rating (CoRa-O). The psychometric properties of the measurements were investigated. The analysis suggested satisfactory psychometric properties of the presented measurements, some limitations are discussed. All measurements are promising instruments for multi-informant assessment of therapeutic competence. They assess the same components of therapeutic competence, so comparisons among the perspectives are possible.
Key words: therapeutic competence, assessment, multi-informant rating.

How to cite this paper: Koddebusch C & Hermann C (2019). Multi-informant Assessment of Therapeutic Competence. Development and Initial Validation of a Set of Measurements. *International Journal of Psychology & Psychological Therapy, 19, 1,* 15-28.

Novelty and Significance

What is already known about the topic?

- Assessment of therapeutic competence should be multi-informant. The multi-informant rating should ideally contain ratings by the therapist, the client and an independent observer.
- Nevertheless, a set of measures for the different perspectives based on the same theoretical approach is still lacking.

What this paper adds?

- This paper adds a set of multi-perspective ratings of therapeutic competence based on the ratings of the therapist, the client and an independent observer.
- All measures are based on the same theoretical model of Cognitive Behavioral Therapies (CBT). Therefore, comparisons between the perspectives are possible.

The investigation of therapeutic competence requires a clear definition, a model of core competences as well as adequate assessment instruments, there is an ongoing effort to find these. Following Kaslow (2004), competence can be defined as "an individual's capability and demonstrated ability to understand and do certain tasks in an appropriate and effective manner consistent with the expectations for a person qualified by education and training in a particular profession or specialty thereof" (p. 775). In addition, Barber, Sharpless, Klostermann, and McCarthy (2007) adapted a definition from the field of medicine (Epstein & Hundert, 2002) to clinical psychology: "competence can be thought of as the judicious application of communication, knowledge, technical skills, clinical reasoning, emotions, values, and contextual understanding for the benefit of the individual […] being served" (p. 494).

---

* Correspondence concerning this article should be addressed to: Christine Koddebusch, Department of Clinical Psychology and Psychotherapy, Justus-Liebig-University, Otto-Behaghel-Strasse 10F, 35394 Giessen, Germany, Email: Christine.Koddebusch@psychol.uni-giessen.de

Focusing on the assessment of therapeutic competence, a multi-method and multi-informant approach is required (Hatcher, Fouad, Grus, Campbell, McCutcheon, & Leahy, 2013; Kaslow, 2004). There is a broad consensus that "no single method is able to provide a comprehensive assessment of all aspects of [...] competence" (Muse & McManus, 2013, p. 495). Accordingly, therapeutic competences should be assessed from three different perspectives: (a) therapists' self-report, (b) clients' rating and (c) observers' rating. Indeed, all three perspectives allow equally valid assessment of therapeutic competence, since each perspective taps different aspects of therapeutic competence (Orlinsky, Grawe, & Parks, 1994). Investigating the associations among the different perspectives ratings, various studies have shown low (e.g. McManus, Rakovshik, Kennerley, Fennell, & Westbrook, 2012) or even no correlation among the different perspectives (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012; Fitzpatrick, Iwakabe, & Stalikas, 2005; Mallinckrodt, 1993; Mathieson, Barnfield, & Beaumont, 2009). These findings underline the importance of multi-perspective assessment.

Currently, several instruments exist to assess therapeutic competences from different perspectives. The most frequently used perspective for assessing therapeutic competences are ratings by external observers. The current 'gold-standard' for observer-based ratings (McManus, Westbrook, Vázquez Montes, Fennell, & Kennerley, 2010) is the Cognitive Therapy Scale (CTS; Young & Beck, 1980) and its revised version CTS-R (Blackburn, James, Milne, Baker, Standart, Garland, & Reichelt 2001). Muse and McManus (2013) summarized the research findings of several studies investigating the CTS and reported good internal consistencies but also high intercorrelations between the subscales and furthermore instable inter-rater reliability. As a main criticism they reported poor content validity because of overly focusing on the treatment of depression and overlapping in the content of items. Focusing on CTS-R, the authors concluded that even this revised version still has the same limitations. For both versions, the authors critically note that the cut-off points are not empirically proved. In addition, Roth (2016) questioned the generic nature of the CTS by pointing out the need to adapt the scale in dependence of the conditions of the therapeutic setting. Recently, Roth (2016) published the University College London-Scale (UCL-Scale) based on the competence framework (Roth & Pilling, 2007). Yet, the evaluation study is still running, so the psychometric quality of the scale is still unknown. Furthermore, Muse, McManus, Rakovshik, and Thwaites (2017) developed the Assessment of Core-CBT-Skills (ACCS) based on the core competences delivered by Roth and Pilling (2007), the CTS (Young & Beck, 1980) and the CTS-R (Blackburn *et alii*, 2001). Beside a self-rating version, the ACCS entails also an assessor version. For both versions, the reported psychometric analyses showed satisfying results.

There are only few self-report measures available for Observers' rating. For example, Bennett-Levy and Beedie (2007) developed the Cognitive Therapy Self-Rating Scale (CTSS). Items of the CTSS resulted of a modification of items of the CTS (Young & Beck, 1980). Despite scores of Cronbachs Alpha were adequate (Bennett-Levy & Beedie, 2007), the CTSS lacks further analysis of reliability and validity. Another self-rating questionnaire focusing on therapists perception of relevant working mechanism of the therapeutic process (Grawe, 2000) is the therapist version of the Berner Post Session Report (TSTB, Flückiger, Regli, Zwahlen, Hostettler, & Caspar, 2010). Its subscales were well confirmed by confirmatory factor analysis, reported reliability and validity were satisfying (Flückiger *et alii*, 2010). While some subscales (e.g. resource activation) may well be used for the assessment of therapeutic competence, others are more specific variables of the therapeutic process (e.g. openness) and, therefore, not applicable to

the assessment of therapeutic competence. All in all, therapists' self-assessment of therapeutic competence is frequently used because of low costs and simple application. However, it must be considered that self-ratings are personally biased (Kaslow, Grus, Campbell, Fouad, Hatcher, & Rodolfa, 2009; Mathieson *et alii*, 2009). Some authors even make a clear recommendation not to use therapists self-ratings for standardized formal assessment of therapeutic competence (Muse & McManus, 2013) and advice to use self-ratings only for self-reflection (e.g. in training or during independent practice).

There are only few instruments assessing clients' rating of therapeutic competence. From the existing ones, most deal with clients' perception of the therapeutic relationship (e.g. Working Alliance Inventory (WAI, Horvath & Greenberg, 1989)). A more general assessment of relevant aspects of the therapeutic process allows the patient version of the Berner Post Session Report (PSTB, Flückiger *et alii*, 2010) with focus on Grawe's working mechanisms (2000). The reported psychometric quality is as satisfying as for the therapists' version TSTB. Like for the TSTB, some scales of the PSTB can be interpreted as therapeutic competence (e.g. experience of clarification) whereas others asses process variables that cannot be interpreted as therapeutic competence (e.g. coping experiences).

Nevertheless, there is an ongoing controversy if clients' perspective is of relevance for assessing therapeutic competence. On the one side, it is called that assessing the client perspective increases the quality of the assessment of therapeutic competence (Lichtenberg, Portnoy, Bebeau, Leigh, Nelson, Rubin, Leon, & Kaslow, 2007). On the other side, it has been argued that clients' view on the therapist's competence is likely not to match what therapists themselves as well as observers would consider to be competent (Fitzpatrick *et alii*, 2005).

Overall, the requirement to consider multi-perspective assessment of therapeutic competence is actually not met satisfactory. Although there are several instruments to assess therapeutic competence from a certain perspective, the assessments are mostly limited to just one perspective. The newly developed ACCS (Muse *et alii*, 2017) incorporates versions for self-rating and for assessor-rating, but does not consider the client perspective. Furthermore, the assessment tools are based on different theoretical foundations, so finally the results are not comparable and measurements assessing different perspectives cannot be integrated (Muse & McManus, 2013). Finally, the existing instruments cannot be used for the assessment of therapeutic competence of novice therapists starting their clinical training because some items require a level of competence above the one that can be expected at an initial stage (e.g. item *application of cognitive therapy techniques* from the CTS (Young & Beck, 1980)).

In sum, there is a lack of multi-informant measurements of therapeutic competence sharing the same theoretical basis, being applicable at initial stages of clinical training, empirically validated and efficient in practical use.

Our set of multi-informant measurements was developed based on theoretical considerations about central components illustrated in the *Three Level Model of Therapeutic Competence* (Koddebusch & Hermann, 2018). This model summarizes currently discussed aspects of therapeutic competence and integrates these aspects on three levels: First, on a rather stable individual level of *Dispositions* (interpersonal style, personality factors, self-esteem, self-regulation, self-reflection and analytical thinking) which are postulated to be the groundwork for the achievement of therapeutic competence. We further distinguish between the second level *Basic Competences* (communicative competence, interpersonal competence and intrapersonal competence), which are universal components that are postulated to be essential for any therapist regardless of his/her theoretical background

and on the third level *Specific Competences* (cognitive behavioral interventions, diagnostic competence, knowledge and case conceptualization) which differ depending on the theoretical orientation (e.g. CBT).

On the basis of the Three Level Model our aim was to develop multi-method measurement tools for assessing therapeutic competences from the perspectives of therapists, clients and observers.

Method

Participants

A total of 46 graduate student therapists (age: *M*= 24.6 years, *SD*= 2.0; 96% female) self-rated their perceived therapeutic competence. Also, a total of 96 student clients (age: *M*= 25.3 years, *SD*= 5.1; 80% female) rated their student therapist.

Materials and Measures

Measures were developed and evaluated as part of a therapeutic training program at the Department of Clinical Psychology at Justus Liebig University Giessen, Germany. The training program was inspired by the work of Miriam Stein (Stein, 2014). Advanced Master level students in clinical psychology were trained in basic therapeutic and certain CBT-skills. They provided individual sessions in stress management to student clients not enrolled in a psychology program. The clients were seeking help for problems related to student life (e.g. time management, relaxation techniques, emotion regulation). Usually, each student therapist provides 10 individually tailored sessions to two clients. Due to the exploratory and innovative orientation of this work, the measures presented in this study are limited to the components *communicative competence* and *interpersonal competence* (Basic Competences), and *cognitive behavioral interventions* (Specific Competence) of the Three Level Model.

Therapist's global rating. To measure the global self-assessment of therapeutic competence we developed the 22-item questionnaire *Global Rating-Therapist* (GloRa-T). GloRa-T consists of two subscales Basic Competences [communication skills (#1-5) and interpersonal competences (#6, 7)] and Specific Competences (cognitive behavioral interventions #8-22). The component cognitive behavioral interventions includes general CBT competences (#8-17) and in addition techniques of Kanfer's (2006) self-management therapy (#18-22). Therapists rate how much they agree on a five point Likert scale ranging from 0= not at all to 4= fully agree. Subscales are formed by summing up the item raw scores and dividing by the number of items per scale.

Therapist's session rating. For assessing self-perceived in-session therapeutic competence, we developed a session rating questionnaire (*Session Rating Therapist*, SeRa-T), developed based on a selection and an adaption of items from the TSTB (Flückiger *et alii*, 2010). In addition, some newly formulated items were added (see Table 3) for those components of the Three Level Model which are not represented by items of the TSTB. SeRa-T consists of 27 items, which assess Basic Competences [communicative competence (#4-6), interpersonal competence (#1-3, 7)], and Specific Competences (cognitive behavioral interventions #8-27). The component cognitive behavioral interventions includes general CBT skills (#8-22) and in addition items assessing the competence in techniques of Kanfer's (2006) self-management therapy (#23-27). According to TSTB therapists rate their agreements to items 1-7 on a bipolar seven point Likert scale (-3= not at all to

+3= yes, exactly). These item answers are for analysis recorded to 0-6. For items 8-27 therapists rate on a five point Likert scale (0= not at all to 4= fully agrees) how much they agree. Subscales are formed by summing up the item raw scores and dividing by the number of items per scale.

Clients' session rating. Similar to the therapists' version, the session rating questionnaire for clients (*Session Rating Client*, SeRa-C) was developed based on some items taken from the PSTB (Flückiger *et alii*, 2010). In addition, two items were taken from the Therapy Rating Sheet by Schindler, Hohenberger-Sieber, and Hahlweg (1990). Furthermore, new items were added (see Table 4) for those components of the Three-Level-Model which were not covered by the PSTB and the Therapy Rating Sheet. SeRa-C contains of 20 items assessing the subscales Basic Competences (communicative competence #7, 14; interpersonal competence #1, 5, 6) and Specific Competences (cognitive behavioral interventions #2-4, 8-13, 15, 16; techniques in self-management therapy #17-20). According to the PSTB, clients rate their agreement on a seven point Likert scale (-3= not at all to +3= yes, exactly). For analysis item answers are recorded to 0-6. Subscales are formed by summing up the item raw scores and dividing by the number of items per scale.

Observer rating. The Competence Rating for Observer (CoRa-O) was used to assess observer ratings of therapeutic competence based on videotapes of sessions. CoRa-O consists of 12 items including two subscales (*Basic Competences* and *Specific Competences*), a global rating of therapeutic competence (# 1) and a rating of the difficulty of working with this specific client (#14). The subscale Basic Competences consists of the components communicative competence (#2-4) and interpersonal competences (#5, 6). The subscale Specific Competences (#7- 13) entails items assessing cognitive behavioral interventions (#7-12) and techniques in self-management (#13) therapy. Observers rate on a five point Likert scale (1= barely competent, 3= moderately competent to 5= very competent) how competent they perceived the therapist. For the end and midpoints of the scale (scale levels 1, 3 and 5) verbal anchors were formulated. Subscales are formed by summing up the item raw scores and dividing the sum by the number of items per scale. For the present study, video-based observer ratings were obtained for N= 71 sessions. Raters were two female psychology students. The raters participated in 15 hours of training. Observers rated student therapists' competence for a window of 20 minutes (minute 20 to 40 of the one-hour sessions). Interclass coefficients (ICC) as measurements of interrater reliability were calculated for the total scale and for individual items. The interclass coefficient for the total scale was ICC= .63, which is below the recommended lower limit of .75 (Portney & Watkins, 2014). For the individual items ICCs were spread over a large range (.41≤ to < .68) and are all just moderate.

Procedure

Student therapists' global self-ratings GloRa-T was conducted via an online platform prior to the beginning of the training and after completing sessions with both clients. Session ratings (SeRa-T, SeRa-C) were completed after sessions 2 and 9 in a paper-pencil version. For evaluating SeRa-T we used the session ratings of the ninth session with the second clients. Three student therapists missed completing these session ratings and three clients dropped out of the therapy so finally 43 surveys were evaluated. For SeRa-C the ratings of all clients of the ninth session were used. Five clients dropped out of the project, six clients had not participated in the study and three clients missed completing session ratings. Finally, session ratings of 82 clients were evaluated (40 first clients and 42 second clients). Sessions 2 and 9 were videotaped.

Psychometric analysis of CoRa-O referred to the records of the ninth session. 11 clients did not agree with the video recording, 5 clients dropped out of the project, 5 tapes must be excluded due to technical problems and 4 videos were used for training

and therefore excluded from the rating. Finally 71 videotapes of session 9 were rated (36 with the first client, 35 with second client).

Data Analysis

Descriptive statistics [Mean (*M*), Standard Deviation (*SD*)] and results of item analysis (minimum and maximum ratings, range, skewness and kurtosis) were described. Item total correlations values above r_{it}= .30 were interpreted as good (Fisseni, 2004). According to Field (2013) skewness and kurtosis were interpreted as follows: The values of skewness and kurtosis were divided by their standard error (*SE*). If the absolute value of the resulting score was greater than 1.96, skewness and kurtosis were interpreted as significantly differing from the normal distribution. Subsequently, at the level of the subscales, means, standard deviations and item total correlations were conducted. We used Cronbach's α as measurement for internal consistency. Interrater reliability of CoRa-O was determined by calculating intraclass-correlations (Shrout & Fleiss, 1979), unadjusted two-way random model, all ratings from both raters (ICC 2,2). All statistical analyses were calculated using SPSS 22.

RESULTS

Means of all GloRa-T items gathered around the midpoint of the Likert scale, *M*= 2.93, *SD*= .77 (#18) to *M*= 3.63, *SD*= .53 (#6), see Table 1. The range of the item scores was generally rather small, with no ratings on the end points of the Likert scale.

Table 1. Items of the GloRa-T (Global Self Rating-Therapists) and scale properties.

		M (*SD*)	min-max	Skewness (*SE*= .35)	Kurtosis (*SE*= .69)
1*	I can meet clients with appreciation.	3.59 (.50)	3-4	-0.37	-1.95[+]
2*	I can feel with the clients and understand their problems.	3.24 (.48)	2-4	0.62	-0.11
3*	I really fulfill the role of being a counselor, I don't just pretend to do so.	3.48 (.55)	2-4	-0.34	-1.05
4*	I can listen attentively.	3.50 (.62)	2-4	-0.86[+]	-0.21
5*	I can establish a warm and welcoming atmosphere.	3.48 (.59)	2-4	-0.60	-0.55
6*	I can built up a relationship with clients.	3.63 (.53)	2-4	-1.02[+]	-0.03
7*	I can take up a friendly, yet professional position towards clients.	3.04 (.67)	1-4	-0.52	1.06
8	I can work goal-oriented.	3.26 (.61)	2-4	-0.20	-0.50
9	I can establish a positive efficacy expectation.	2.96 (.70)	1-4	-0.35	0.31
10	I can refer to the client's current life circumstances.	3.26 (.65)	2-4	-0.31	-0.64
11	I can notice the clients' resource and strengths.	3.26 (.71)	2-4	-0.43	-0.90
12	I can develop strategies for problem-solving.	3.13 (.65)	2-4	-0.14	-0.58
13	I can emphasize on new approaches to various problem areas and their relationship.	3.17 (.64)	1-4	-0.69[+]	1.85[+]
14	I can motivate clients to participate actively.	3.15 (.73)	1-4	-0.60	0.36
15	If it is necessary and reasonable, I can let go of my pre-structured session agenda and attend to the clients' recent situations.	3.43 (.69)	2-4	-0.83[+]	-0.46
16	I can structure the sessions.	3.50 (.51)	3-4	0.00	-2.09[+]
17	I can choose reasonable interventions.	3.24 (.64)	2-4	-0.25	-0.58
18	I can turn the clients' focus towards controllable behavioral variables. and away from personality traits.	2.93 (.77)	1-4	-0.19	-0.54
19	I can point out a reasonably positive view to the clients.	3.20 (.58)	2-4	-0.03	-0.19
20	In cooperation with the client I can divide the relevant problem into sub-steps.	3.22 (.73)	2-4	-0.36	-1.00
21	I can encourage clients to let go of fixed opinions and provide flexible viewpoints instead.	3.11(.67)	2-4	-0.13	-0.72
22	I can help the clients to focus on the future.	3.20 (.72)	2-4	-0.31	-0.97

Notes: The original items were in German (German version is available from the corresponding author); *= Items a priori postulated as assessing *Basic Competences*; [+]= Values of skewness and kurtosis significantly differing from the normal distribution; *M*= Mean; *SD*= Standard Deviation; *SE*= Standard Error.

Furthermore, the items were skewed to the left, analysis of kurtosis showed that most items were leptokurtic. Regarding the distribution, items were almost all negatively skewed to the right with frequent higher ratings and mostly negative kurtosis (platykurtic) with fewer score at the ends. All in all, the distributions of the items were tolerable with exceptions of items 1 (Kurtosis= -1.95, *SE*= .69), 4 (Skewness= -0.86, *SE*= .35), 6 (Skewness= -1.02), 13 (Skewness= -0.69, Kurtosis= 1.85), 15 (Skewness= -0.83) and 16 (Kurtosis= -2.09) that differ significantly from normal distribution.

The subscales showed good or acceptable internal consistencies (α= ≤72 to <.86), except for subscale interpersonal competences (α= .29). Therefore, the two items of the subscale interpersonal competence (working alliance #6 and role behavior #7) were considered separately in former analysis and no item discrimination coefficients were calculated. Accordingly, for this scales internal consistencies and item discrimination coefficients were not calculated. The remaining item discrimination coefficients of the subscales were all above the recommended lower limit of r_{it}= .3 (See Table 2). GloRa-T's subscales communicative competence, working alliance, CBT interventions and techniques in SMT were highly correlated (r_{xy}= ≤44 to <.86). In contrast, there were no correlations between role behavior and the other subscales.

Table 2. Subscales of the measurements of therapeutic competence.

			M (SD)	r_{it} (min/max)	α
GloRa-T	Basic competences	Communicative competence	3.46 (.38)	.30/.62	.72
		Working alliance	3.63 (.53)	-	-
		Role behavior	3.04 (.67)	-	-
	Specific competences	Cognitive behavioral interventions	3.24 (.44)	.46/.66	.86
		Techniques in SMT	3.13 (.50)	.39/.61	.76
SeRa-T	Basic competences	Communicative competence	5.29 (.58)	.30/.53	.57
		Working alliance	5.10 (.62)	.43/.59	.69
		Role behavior	5.02 (.80)	-	-
	Specific competences	Cognitive behavioral interventions	3.06 (.52)	.23/.76	.88
		Techniques in SMT	2.96 (.66)	.57/.65	.82
SeRa-C	Basic competences	Communicative competence	5.14 (.73)	.56	.72
		Interpersonal competence	5.36 (.62)	.39/.51	.64
	Specific competences	Cognitive behavioral interventions	4.77 (.67)	.32/.57	.87
		Techniques in SMT	4.90 (.69)	.32/.57	.68
CORA-O	Basic competences	Communicative competence	2.77 (.63)	.61/.83	.85
		Interpersonal competence	2.92 (.84)	.73	.84
	Specific competences	Cognitive behavioral interventions	3.00 (.58)	.56/.73	.87
		Techniques in SMT	2.68 (.62)	-	-

Notes: *M*= Mean; *SD*= Standard Deviation; r_{it}= minimal and maximal Item Discrimination Coefficients; α= Internal Consistency (Cronbach's α).

Session-Rating-Therapist (SeRa-T). For the first part for the SeRa-T (#1-7), averaged means were almost in the upper half of the scale, *M*= 4.95, *SD*= .90 (#3) to *M*= 5.44, *SD*= .63 (#4). In the second part (#8-27) means were also above the middle point *M*= 2.23, *SD*= 1.00;(#11) and *M*= 3.47, *SD*= .63 (#18), see Table 3. Range was rather small. All in all, items have a negatively skewed distribution with many high scores and kurtosis was mostly positive (leptokurtic) with many scores in the tails of the distribution. However, several items (1, 5, 6, 8, 14, 16-20, 22-24) had a critical skewness (≤-3.53 to <-.75, *SE*= .36) and kurtosis (#6, 12, 14, 17, 19, 20, 22, 23; ≤2.02 to <17.98, *SE*= .71) and differed significantly from the normal distribution.

Table 3. Items of the SeRa-T (Session Rating-Therapists) and scale properties.

	M (SD)	min-max	Skewness (*SE*= .36)	Kurtosis (*SE*= .71)
1* Today I felt comfort in my relationship with the client.[B]	5.14 (.80)	3-6	-0.84[+]	0.62
2* The client and me understood each other.[B]	5.21 (.64)	4-6	-0.21	-0.55
3* The client and me work on mutual goals.[B]	4.95 (.90)	3-6	-0.53	-0.41
4* I met my client with appreciation.[N]	5.44 (.63)	4-6	-0.68	-0.45
5* I empathized with the client and understood his problems.[N]	5.21 (.71)	3-6	-0.75[+]	0.93
6* I really fulfilled my role as a counselor and did not only "played" it.[N]	5.23 (1.00)	0-6	-3.53[+]	17.98[+]
7* I took up a friendly but not too private attitude towards the client.[N]	5.02 (.80)	3-6	-0.62	0.24
8 Today I tried to use the client's strengthen in a targeted manner.[B]	2.95 (.87)	1-4	-0.81[+]	0.37
9 Today I worked towards a better client's handling of situations that are difficult for him.[B]	3.40 (.62)	2-4	-0.51	-0.57
10 Today I tried to improve the client's action competence in a targeted manner[B]	3.12 (.76)	1-4	-0.54	-0.02
11 Today I have touched on sore spots of the client.[B]	2.23 (1.00)	0-4	0.11	-0.49
12 Today I used the opportunity that the client experiences his positive sides.[B]	3.02 (.94)	0-4	-1.14	1.61[+]
13 Today I tried that the client feels more self confident for the solution of problems.[B]	3.16 (.72)	1-4	-0.66	0.62
14 Today I tried intensively to improve clients' worth.[B]	3.14 (.94)	0-4	-1.19[+]	1.68[+]
15 Today I tried that the client sees his problems in new contexts.[B]	2.52 (1.04)	0-4	-0.27	-0.53
16 Today I tried to refer to clients current life circumstances.[B]	3.21 (.83)	1-4	-0.94[+]	0.51
17 I tried that the client accepts responsibility.[N]	2.86 (.89)	0-4	-1.00[+]	1.64[+]
18 Today I worked goal-oriented.[N]	3.47 (.63)	2-4	-0.76[+]	-0.36
19 I conveyed to the client that the counseling might have positive influence on his problems.[N]	3.19 (.91)	0-4	-1.39[+]	2.62[+]
20 I tried that the client works on his own initiative.[N]	3.21 (.74)	1-4	-1.10[+]	2.02[+]
21 I adapted my agenda to the current needs of the client.[N]	3.33 (.64)	2-4	-0.42	-0.63
22 Today I have actively tried to direct clients attention on influenceable behavior.[N]	3.09 (.84)	0-4	-1.19[+]	2.90[+]
23 I encouraged the client to a solution-oriented approach.[N]	3.23 (.72)	1-4	-1.19[+]	2.68[+]
24 I have tried to offer a positive perspective to the client.[N]	3.33 (.75)	1-4	-0.98[+]	0.82
25 I tried to show the client to divide problems into sub-steps.[N]	2.65 (1.04)	0-4	-0.42	-0.40
26 I have tried to convex to the client flexible attitudes toward different solutions.[N]	2.49 (1.22)	0-4	-0.22	-1.01
27 I directed the view of the client on the future.[N]	2.95 (.75)	1-4	-0.27	0.27

Notes: The original items were in German (German version is available from the corresponding author); *= Items a priori postulated as assessing Basic Competences, ratings of these items were done on a Likert scale from (-3) to (+3), recorded as 0 to 6, the remaining items were rated on a Likert-Scale from 0 to 4; [B]= Items from the TSTB (Flückiger *et alii*, 2010); [N]= New formulated item. [+]= Values of skewness and kurtosis significantly differing from the normal distribution; *M*= Mean; *SD*= Standard Deviation; *SE*= Standard Error.

For reasons of consistency the subscale interpersonal competence was also split into two further subscales working alliance (#1, 2, 3) and role behavior (#7). Internal consistencies of all subscales ranged from critical to good (α= ≤57 to <.88) (see Table 2). Item discrimination coefficients were good and above the recommended level, except for item 11 (subscale CBT interventions; r_{it}= .23). Almost all subscales were significantly intercorrelated (r_{xy}= ≤.32 to <.83) except correlations between communicative competence and techniques in SMT that reached only a borderline significance (r= .29+) and between working alliance and role behavior (r_{xy}= .11).

Client's session rating (SeRa-C). Averaged means of SeRa-C were mostly in the upper half of the Likert skale, *M*= 3.55, SD= 1.25 (#15) to *M*= 5.41, *SD*= .63 (#17), see Table 4. Distributions were skewed right with the frequent scores clustered at the higher end and rather positive kurtosis (leptokurtic). Skewness of almost all items different significantly from normal distribution (skewness= ≤-2.01 to <.75, *SE*= .27), kurtosis of half of the items differed significantly from normal distribution (kurtosis= ≤-1.21 to <8.47, *SE*= .53).

Subscales communicative competence and CBT interventions had acceptable to good internal consistencies (α= .72 and .87) (see Table 2). In contrast the subscales interpersonal competences and techniques in SMT had poor internal consistencies (α= .64 and .68). Item discrimination coefficients of the subscales were all above the

Table 4. Items of the SeRa-C (Session Rating-Clients) and scale properties.

		M (SD)	min-max	Skewness (SE= .27)	Kurtosis (SE= .53)
1*	Today I was comfortable with my counselor. [B]	5.40 (.89)	2-6	-1.87[+]	4.09[+]
2	I feel that I can better understand me and my problems.[B]	5.21 (.83)	3-6	-0.81[+]	0.02
3	Today we got closer to the core of my problems.[B]	4.85 (.89)	3-6	-0.14	-0.97
4	The counselor shows me my strengths.[B]	5.20 (.92)	3-6	-0.89[+]	-0.19
5*	My counselor and I understand each other.[B]	5.40 (.73)	3-6	-1.18[+]	1.27[+]
6*	I think my counselor really cares about my wellbeing.[B]	5.28 (.81)	1-6	-2.01[+]	8.47[+]
7*	Today I felt understood from my counselor.[S]	5.24 (.81)	2-6	-1.34[+]	2.74[+]
8	I experienced the session as planned and purposeful.[S]	5.34 (.72)	3-6	-1.03[+]	1.11[+]
9	At the moment my counselor supports me in how I would like to be.[B]	5.23 (.86)	3-6	-0.94[+]	0.16
10	I fell more confident that I can solve my problems on my own.[B]	4.93 (1.07)	1-6	-0.95[+]	1.11[+]
11	Now I know better what I want.[B]	4.56 (1.15)	1-6	-0.81[+]	0.80
12	Today I was emotionally heavily involved.[B]	3.89 (1.33)	1-6	-0.05	-0.75
13	Today the counselor directed my attention to behaviors that I can actively influence.[N]	5.05 (.84)	1-6	-0.75[+]	0.26
14*	I feel that my counselor appreciates me.[B]	5.04 (.85)	1-6	-1.66[+]	5.60[+]
15	I was strongly affected by what we have done today.[B]	3.55 (1.25)	0-6	0.06	0.18
16	I get along with situations better than before.[B]	4.74 (1.00)	3-6	-0.06	-1.21[+]
17	Today the counselor was bothered that I get a positive perspective.[N]	5.41 (.63)	3-6	-0.90[+]	1.32[+]
18	Today I experienced that my problem consists of several small parts.[N]	4.60 (1.21)	0-6	-0.99[+]	1.32[+]
19	Today the counselor conveyed to me that there are usually several ways to solve a problem.[N]	4.59 (1.03)	2-6	-0.48	-0.41
20	Today the counselor directed my attention to the future.[N]	5.02 (.93)	3-6	-0.81[+]	-0.07

Notes: The original items were in German (German version is available from the corresponding author); *= Items a priori postulated as assessing Basic Competences; [B]= Items from the TSTB (Flückiger *et alii*, 2010); [N]= New formulated item; [S]= Items extracted from the SB-K (Schindler *et alii*, 1990); [+]= Values of skewness and kurtosis significantly differing from the normal distribution; *M*= Mean; *SD*= Standard Deviation; *SE*= Standard Error.

recommended level (r_{it}= ≤32 to <.57). All subscales had very high intercorrelations (r_{xy}= ≤.46 to <.74).

Observer rating (CoRa-O). Means of observer ratings were slightly below or almost around the Likert's scales midpoint, *M*= 1.85, *SD*= .62 (#2) to *M*= 3.37, *SD*= .83 (#12), see Table 5. Global competence rating (#1) was also around the Likert scale's midpoint (*M*= 2.88, *SD*= .65). The difficulty of the client (# 14) was rated as moderately (*M*= 2.40, *SD*= .86). Regarding the distribution, most items were negatively skewed with frequent scores at the higher end of the scale and in addition, some items were leptokurtic others platykurtic. However, skewness of items 1, 3, 4, 7, 9, 11 and 12 (skewness= ≤-0.81 to <-0.56, *SE*= .29) and kurtosis of item 4 (kurtosis= 1.22, *SE*= .57) differed significantly from normal distribution.

Table 5. Items of the CoRa-O (Competence Rating-Observer) and scale properties.

		M (SD)	min-max	Skewness	Kurtosis	ICC(2,2)
1	Global competence	2.88 (.65)	1-4	-0.56	0.84	.54***
2*	Empathy	1.85 (.62)	1-3.5	0.31	-0.36	.53***
3*	Basic attitude	3.20 (.83)	1-4.5	-0.74	0.22	.66***
4*	Communication skills	3.26 (.71)	1-4.5	-0.81	1.22	.60***
5*	Working alliance	2.67 (.73)	1-4	-0.53	-0.19	.49***
6*	Role behaviour	3.18 (.85)	1-5	-0.47	0.46	.64***
7	Solution orientation	2.84 (.70)	1-4	-0.61	0.37	.51***
8	Resource orientation	2.23 (.80)	1-4	0.35	-0.43	.57***
9	Encourages active engagement of the client	3.07 (.67)	1-4	-0.66	0.61	.46***
10	Positive efficacy expectation	3.13 (.72)	1.5-4.5	-0.40	-0.51	.43***
11	Flexibility	3.11 (.71)	1-4	-0.66	0.21	.48***
12	Structuring	3.37 (.83)	1-5	-0.77	0.74	.68***
13	Encouragement of clients self-management	2.68 (.62)	1-4	-0.24	-0.01	.41***
14	Difficulty of the client	2.40 (.86)	1-5	0.54	0.25	.52***

Notes: The original items were in German (German version is available from the corresponding author); *= Items a priori postulated as assessing *Basic Competences*; ***= *p* ≤ .001.

Internal consistency of all subscales was good (α= ≤.84 to <.87), see Table 2; all item discrimination coefficients were good (r_{it}= ≤.56 to <.83) and all subscales were very highly intercorrelated (r_{xy}= ≤.66 to <.90).

Intercorrelations of the multi-informant ratings. When comparing therapists' global and in-session self-ratings (GloRa-T and SeRa-T), only subscales communicative competence (r_{xy}= .35; p <.05) and CBT-interventions (rxy= .50; p < .01) were significantly correlated.

Looking at the multi-informant ratings, results show that ratings of therapists, clients and observers were correlated significantly only in some isolated components of therapeutic competence. Comparing the different perspectives, in-session ratings of therapists' and clients' were only correlated for therapists' ratings of working alliance (SeRa-T) and clients' ratings of interpersonal competence (SeRa-T) (r_{xy}= .37; p <.05). Comparing therapists' and observers' session ratings, only competence ratings of CBT interventions were significant correlated (r_{xy}= .48; p ≤.05). Regarding observers' and clients' ratings only their ratings of communicative competence were borderline significantly correlated (r_{xy}= .31; p= .08).

Discussion

The present study introduced and evaluated a set of measurements of therapeutic competence that are based on the Three-Level Model of Therapeutic Competence (Koddebusch & Hermann, 2018) and allow multi-informant assessments of therapists', clients' and observers' ratings. This set allows direct comparison of the different rating perspectives because all measurements entail the same components of therapeutic competence.

Therapists' global self-ratings of their therapeutic competence (GloRa-T) were comparably high for all items as reflected in high means, small ranges and negative skewed distributions. Furthermore, the distribution of several items (skewness and kurtosis) differed significantly from normal distribution. In general, regarding the distribution of GloRa-T as well as concerning the other measurements, it is a question whether one can ever expect a normal distribution. This applies to any therapeutic sample but particularly for the sample of this study that is high self-selected because student therapists have actively chosen to participate in the training. However, internal consistencies of the subscales communicative competence, CBT interventions and techniques in SMT were good or acceptable. Furthermore item discrimination coefficients of these subscales were adequate. Due to problematic internal consistency and item discrimination coefficients of the subscale interpersonal competences, the two items of this scale were examined separately. This procedure is confirmed when looking at the content of the two items: while item 6 assesses the competence to build and maintain a working alliance (I can build up a relationship with clients), item 7 reflects the competence to an adequate role behavior (I can take a friendly, yet professional position toward clients). Further research on interpersonal competence should consider critically whether building and maintaining of a working alliance and adequate role behavior can be summarized in one subscale.

All in all, there were first indications that GloRa-T is a valid measurement of self-rated therapeutic competence and that a meaningful building of subscales is possible.

The second evaluated self-rating instrument was the session rating SeRa-T. Therapists' in-session ratings were in all items rather high. Several items showed a distribution with skewness and kurtosis significantly differing from normal distribution. Interestingly, among the critical items there are also some items which have been adapted

from the TSTB (Flückiger *et alii*, 2010). These deviations from the normal distribution also in well-evaluated items supports the prior consideration that normal distribution cannot be expected generally in every context.

For reasons of consistency with GloRa-T, in SeRa-T the subscale interpersonal competence was also interpreted separately for the items working alliance and role behavior. Internal consistencies of the subscales of SeRa-T ranged from bad for communicative competence (α= .57) to good for CBT interventions (α= .88). All item discrimination coefficients were above the recommended level of .3 except for item 11 (r_{it}= .23).

Considering the session ratings of clients, analyses of SeRa-C consistently showed relatively high ratings with item means high above the middle point of the Likert Scale. Regarding the distribution, skewness and kurtosis of several items differed significantly from normal distribution, these included both adapted and newly formulated items. With regard to psychometric properties of the subscales, internal consistencies and item discrimination coefficients of communicative competence and CBT interventions were satisfying. The subscales techniques in SMT and interpersonal competences, however, had lower internal consistencies, yet item discrimination coefficients were tolerable. All subscales were highly intercorrelated.

With regard to CoRa-O, descriptive inspection showed ratings around the midpoint of the Likert-Scale. Regarding the distributions of the items, some differ significantly from normal distribution, especially in skewness. Internal consistencies and item discrimination coefficients of the subscales were acceptable. All subscales were highly intercorrelated. The low inter-rater reliability (ICC= .63), however, was problematic. Possibly, the low agreement was a consequence of our raters being novices. Weck, Hilling, Schermelleh-Engel, Rudari, and Stangier (2011) compared competence ratings of novice and expert raters and found that ratings of novice raters were less reliable than those of expert raters. Due to these findings further analysis of CoRa-O with more experienced raters are necessary.

In all four investigated measurements the items reported as critical (e.g. those that significantly differ from the normal distribution or with unsatisfactory values of internal consistencies) need further consideration. However, they should not yet be eliminated from the scales due to the exploratory character of these analyses and the relatively small sample size. In addition, the former consideration that usual psychometrical claims regarding normality are not fully adaptable to this kind of assessment represents a further reason to keep the items in the questionnaires. Furthermore, during the initial developmental stage of measurement tools even low levels of internal consistency (i.e. α <.5) are tolerable (Field, 2013). Nevertheless, further psychometric analyses based on greater sample sizes allowing for factor analysis need to carefully address the items in question.

Considering all questionnaires, the subscales communicative competence, working alliance, CBT interventions and techniques in SMT were highly correlated. In contrast, the subscales role-behavior was not correlated with any other subscales. These findings are consistent with those of other measurements of therapeutic competence. For example, Weck, Hautzinger, Heidenreich, and Stangier (2010) reported a correlation of .59 (p ≤.001) for the two subscales of the German Version of the CTS. These high intercorrelations led to the consideration that there is maybe only one global factor of therapeutic competence. Accordingly, the subdivision into several individual components is not necessary or neither possible. In this case, the formation and further interpretation of total scores would be quite acceptable.

When investigating the relationship between the different perspectives by analyzing the intercorrelations of the measurements, only few perspective/subscales were correlated. That therapists and clients agree in their rating of working alliance is in line with Mallinckrodt (1993), who reported a positive but small agreement between counselors' and clients' ratings of alliance but in contrast with Fitzpatrick *et alii* (2005), who reported a divergence between the two alliance ratings. Results regarding therapists' in-session ratings (SeRa-T) and observers' ratings (CoRa-O) also revealed very few significant associations as only the ratings of CBT interventions were statistically significant (r_{xy}= .48). This behavioral based competence was possibly easier to observe, conceptualize and in consequence easier to rate than other components of therapeutic competence. Accordingly, this makes it possibly easier for therapists and observer to share a similar conceptualization of the competence CBT interventions. All in all, findings are consistent with previous findings from Mathieson et alii (2009) who also reported no correlations between therapists' self-assessment of therapeutic competence and the assessment by independent observer or supervisors. Rearding observer (CoRa-O) and clients' ratings (SeRa-C), the highest correlation was found for communicative competence (r_{xy}= .31;, *p*= .08). Clients seem to have an intuitive concept of communication competence that matches with the professional concept underlying the observers' ratings. This is another indication that clients' ratings of therapeutic competence are more profound than assumed.

It is interesting that from all perspectives therapeutic competence of the student therapists was rated as quite high. These findings may have been a result of overestimation by the novice therapists which has already been found by Brosan, Reynolds, and Moore (2008). Another possible explanation might be that the sample of our student therapists was highly self-selected. Since the project was associated with a higher workload than common classes, only particularly motivated students might have registered. In addition, providing 10 self-directed sessions to two fellow students may have been deterrent for students less self-confident with regard to their competences. Mallinckrodt & Nelson (1991) supposed that clients' ratings could be biased by knowledge of their own level of expertise as therapists. Accordingly, student clients might have rated their student therapists' competence more positively since they had been informed about all student therapists were beginners. High competence ratings from observers (Cora-O) may be due to observers in the present study being novices. Weck *et alii* (2011) showed that therapeutic competence could not be evaluated satisfying by novice raters. In addition, student raters might be biased when having to rate fellow-students of the same age.

Some limitations of the present study need to be addressed. First, it should be noted, that the study sample was quite small for psychometric evaluations so that certain statistical analyses such as factor analyses could not be calculated. It is important to replicate and expand the findings of the present study. Furthermore, the generalization of the results is limited since our sample consisted of novice student therapist. Although valid measurements of therapeutic competence should also be applicable to the special group of therapist at the beginners' level, further analysis need to include also experienced therapists. Not yet addressed has been the question whether the measurements are sensitive to changes in competence.

Despite some limitations, the initial evaluation suggested satisfactory psychometric properties of the presented measurements. Hence, they are promising instruments for multi-informant assessment of therapeutic competence. Our research has several clinical implications. Having a set of multi-informant measurements may help to close the gap of previously not available multi-informant measurements of therapeutic competence.

Since all measurements were based on the model, ratings of the individual perspectives are comparable.

References

Barber JP, Sharpless BA, Klostermann S, & McCarthy KS (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice, 38*, 493-500. Doi= 10.1037/0735-7028.38.5.493

Bennett-Levy J & Beedie A (2007). The Ups and Downs of Cognitive Therapy Training: What Happens to Trainees' Perception of their Competence During a Cognitive Therapy Training Course? *Behavioural and Cognitive Psychotherapy, 35*, 61-75. Doi= 10.1017/S1352465806003110

Blackburn I-M, James IA, Milne DL, Baker C, Standart S, Garland A, & Reichelt FK (2001). The revised cognitive therapy scale (CTS-R): Psychometric properties. *Behavioural and Cognitive Psychotherapy, 29*, 431-446. Doi= 10.1017/S1352465801004040

Brosan L, Reynolds S, & Moore RG (2008). Self-Evaluation of Cognitive Therapy Performance: Do Therapists Know How Competent They Are? *Behavioural and Cognitive Psychotherapy, 36*, 581-587. Doi= 10.1017/S1352465808004438

Dennhag I, Gibbons MBC, Barber JP, Gallop R, & Crits-Christoph P (2012). Do supervisors and independent judges agree on evaluations of therapist adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research, 22*, 720-730. Doi= 10.1080/10503307.2012.716528

Epstein RM & Hundert EM (2002). Defining and Assessing Professional Competence. *Journal of the American Medical Association, 228*, 226-235.

Field AP (2013). *Discovering statistics using IBM SPSS statistics* (4th edition). SAGE Publications Ltd: London.

Fisseni HJ (2004). *Lehrbuch der psychologischen Diagnostik: Mit Hinweisen zur Intervention* (3., überarb. und erw. Aufl.). Göttingen: Hogrefe.

Fitzpatrick MR, Iwakabe S, & Stalikas A (2005). Perspective divergence in the working alliance. *Psychotherapy Research, 15*, 69-80. Doi= 10.1080/10503300512331327056

Flückiger C, Regli D, Zwahlen D, Hostettler S, & Caspar, F. (2010). Der Berner Patienten- und Therapeutenstundenbogen 2000. *Zeitschrift für Klinische Psychologie und Psychotherapie, 39*, 71-79. Doi= 10.1026/1616-3443/a000015

Grawe K (2000). *Psychologische Therapie* (2. korrigierte Aufl.). Göttingen: Hogrefe.

Hatcher RL, Fouad NA, Grus CL, Campbell LF, McCutcheon SR, & Leahy KL (2013). Competency benchmarks: Practical steps toward a culture of competence. *Training and Education in Professional Psychology, 7*, 84-91. Doi= 10.1037/a0029401

Horvath AO & Greenberg LS (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*, 223-233. Doi= 10.1037/0022-0167.36.2.223

Kanfer, F. H. (2006). Selbstmanagement- Therapie: Ein Lehrbuch für die Klinische Praxis. Berlin: Springer.

Kaslow, N. J. (2004). Competencies in Professional Psychology. American Psychologist, 59, 774–781.

Kaslow, N. J., Grus, C. L., Campbell, L. F., Fouad, N. A., Hatcher, R. L., & Rodolfa, E. R. (2009). Competency Assessment Toolkit for professional psychology. Training and Education in Professional Psychology, 3(4, Suppl), 27–45. Doi= 10.1037/a0015833

Koddebusch C & Hermann C (2018). A Proposed Conceptualization of Therapeutic Competence: The Three Level Model. Research in Psychotherapy: *Psychopathology, Process and Outcome, 21*, 12-23.

Lichtenberg JW, Portnoy SM, Bebeau MJ, Leigh IW, Nelson PD, Rubin NJ, Leon LS, & Kaslow NJ (2007). Challenges to the assessment of competence and competencies. *Professional Psychology: Research and Practice, 38*, 474-478. Doi: 10.1037/0735-7028.38.5.474

Mallinckrodt B (1993). Session Impact, Working Alliance and treatment outcome in brief counseling. *Journal of Counseling Psychology, 40*, 25-32.

Mallinckrodt B & Nelson NL (1991). Counselor Training Level and the Formation of the Psychotherapeutic Working Alliance. *Journal or Counseling Psychology, 38*, 133-138.

Mathieson FM, Barnfield T, & Beaumont G (2009). Are we as good as we think we are? Self-assessment versus other forms of assessment of competence in psychotherapy. *The Cognitive Behaviour Therapist, 2*, 43-50.

Doi: 10.1017/S1754470X08000081

McManus F, Rakovshik S, Kennerley H, Fennell M, & Westbrook D (2012). An investigation of the accuracy of therapists' self-assessment of cognitive-behaviour therapy skills. *British Journal of Clinical Psychology, 51*, 292-306. Doi: 10.1111/j.2044-8260.2011.02028.x

McManus F, Westbrook D, Vázquez Montes M, Fennell M, & Kennerley H (2010). An evaluation of the effectiveness of Diploma-level training in cognitive behaviour therapy. *Behaviour Research and Therapy, 48*, 1123-1132. Doi: 10.1016/j.brat.2010.08.002

Muse K & McManus F (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review, 33*, 484-499.

Muse K, McManus F, Rakovshik S, & Thwaites R (2017). Development and psychometric evaluation of the Assessment of Core CBT Skills (ACCS): An observation-based tool for assessing cognitive behavioral therapy competence. *Psychological Assessment, 29*, 542-555. Doi: 10.1037/pas0000372

Orlinsky DE, Grawe K, & Parks BK (1994). Process and outcome in psychotherapy-noch einmal. In AE Bergin & SL Garfield (Eds.), *Handbook of Psychotherapy and Behavior Change*, 4th Ed (pp. 270-376). New York: Wiley.

Portney LG & Watkins MP (2015). *Foundations of clinical research: Applications to practice*, 3rd Ed: Upper Saddle River, NJ: Pearson/Prentice Hall.

Roth AD (2016). A New Scale for the Assessment of Competences in Cognitive and Behavioural Therapy. *Behavioural and Cognitive Psychotherapy, 44*, 620-624. Doi: 10.1017/S1352465816000011

Roth AD & Pilling S (2007). The competences required to deliver effective cognitive and behavioural therapy for people with depression and with anxiety disorders. Retrieved from http://www.ucl.ac.uk/clinical-psychology/CORE/CBT_Competences/CBT_Competence_List.pdf

Schindler L, Hohenberger-Sieber E, & Hahlweg K (1990). *Stundenbeurteilung Klient/Therapeut*. *In Diagnostische Verfahren für Berater*. Weinheim: Beltz Test.

Shrout PE & Fleiss JL (1979). Interclass correlation: Uses in Assessing Rater Reliability. *Psychological Bulletin, 36*, 420-428.

Stein M (2014). *Entwicklung, Implementierung und Evaluation eines verhaltenstherapeutischen Beratungsangebots*. *Coaching für Studierende von Studierenden*. Master Dissertation, Ruprecht-Karls-Universität, Heidelberg. Online available: https://katalog.ub.uni-heidelberg.de/titel/68127936.

Weck, F., Hautzinger, M., Heidenreich, T., & Stangier, U. (2010). Erfassung psychotherapeutischer Kompetenzen. Zeitschrift für Klinische Psychologie und Psychotherapie, 39(4), 244–250. Doi: 10.1026/1616-3443/a000055

Weck F, Hilling C, Schermelleh-Engel K, Rudari V, & Stangier U (2011). Reliability of Adherence and Competence Assessment in Cognitive Behavioral Therapy: Influence of Clinical Experience. *The Journal of Nervous and Mental Disease, 199*, 276-279. Doi: 10.1097/NMD.0b013e3182124617

Young J & Beck AT (1980). *Cognitive Therapy Scale rating manual: Unpublished manuscript*. University of Pennsylvania: Philadelphia, PA. Retrieved from http://members.academyofct.org/files/documentlibrary/CTRS_Manual.pdf.