



Introduction and evaluation of a therapeutic adherence and competence scale for grief-focused cognitive behavioural therapy

Julia Haneveld ¹, Rita Rosner ¹, Anna Vogel ¹, Svenja Mäkitalo ¹, Julia Tremml ², Regina Steil ³, Winfried Rief ⁴ and Hannah Comtesse ¹

¹Department of Psychology, Catholic University Eichstaett-Ingolstadt, Ingolstadt, Germany; ²Department für Psychische Gesundheit, Klinik und Poliklinik für Psychosomatische Medizin und Psychotherapie, Universitätsklinikum Leipzig, Leipzig, Germany; ³Goethe Universität Frankfurt, Frankfurt, Germany; ⁴Universität Marburg, Marburg, Germany

ABSTRACT

Background: There is no therapeutic competence and adherence scale for grief-focused cognitive behavioural therapy (grief-focused CBT). However, given the growing body of evidence for the efficacy of grief-focused CBT, such a scale is needed both to ensure the internal validity of clinical trials and to facilitate psychotherapy process research.

Objective: To develop and undertake a psychometric evaluation of a therapeutic adherence and competence scale for grief-focused CBT.

Method: The scale was developed in two steps. (I) Five experts on the treatment of prolonged grief disorder provided feedback on the relevance and appropriateness of the items. The scale was revised to reflect their feedback. The final therapeutic adherence and competence scale for grief (TACs-G) consisted of 15 adherence and 16 competence items. (II) Psychometric evaluation of the TACs-G was based on the rating of 48 randomly selected PG-CBT sessions by two independent raters. The videos were recorded in the context of a randomized controlled trial (RCT; DRKS00012317.) ICC was used to calculate inter-rater reliability and TACs-G stability over time (re-evaluation of 10 sessions after 12 months).

Results: The five experts confirmed the relevance and appropriateness of the items. Interrater reliability was found to be high for the total adherence and competence scores (ICC = 0.889 and 0.782, respectively) and moderate to excellent for individual items (ICC = 0.509–1.00). The TACs-G stability over time was found to be strong for both adherence (ICC = 0.970) and competence total scores (ICC = 0.965).

Conclusions: The TACs-G for CBT is a reliable instrument that can be used not only to ensure internal validity but is also suited for psychotherapy process studies. Additionally, it provides a valuable database for targeted feedback in training settings.

Introducción y evaluación de una escala de competencia y adherencia terapéutica para la terapia cognitivo conductual centrada en el duelo

Antecedentes: No existe ninguna escala de competencia terapéutica y adherencia para la terapia cognitivo conductual centrada en el duelo. Sin embargo, dado el emergente cuerpo de evidencia para la eficacia de la TCC centrada en el duelo, se necesita una escala para asegurar la validez interna de los ensayos clínicos y para facilitar el proceso de investigación en psicoterapia.

Objetivo: Desarrollar y ejecutar una evaluación psicométrica de una escala de adherencia terapéutica y competencia para terapia cognitivo conductual focalizada en el duelo (PG-CBT por sus siglas en inglés).

Método: La escala fue desarrollada en dos pasos. (I) Cinco expertos en el tratamiento del trastorno por duelo prolongado dieron retroalimentación en la relevancia y la idoneidad de los ítems. La escala fue revisada para reflejar su retroalimentación. La escala final de adherencia y competencia terapéuticas para el duelo (TACs-G por sus siglas en inglés) consiste en 15 ítems de adherencia y 16 ítems de competencia. (II) Se realizó una evaluación psicométrica de la TACs-G basada en el puntaje de 48 sesiones de PG-CBT seleccionadas al azar por 2 evaluadores independientes. Los videos fueron grabados en el contexto de un ensayo controlado aleatorizado (RCT; DRKS00012317.) Se utilizó un coeficiente de correlación intraclase (ICC por sus siglas en inglés) para calcular la confiabilidad inter evaluador y la estabilidad de la TACs-G durante el tiempo (reevaluación de 10 sesiones después de 12 meses).

Resultados: Los cinco expertos confirmaron la relevancia y adecuación de los ítems. La confiabilidad entre evaluadores fue alta para las puntuaciones totales de adherencia y competencia (ICC = 0.889 y 0.782, respectivamente) y de moderada a excelente para ítems

ARTICLE HISTORY

Received 26 October 2021

Revised 4 May 2022

Accepted 12 May 2022

KEYWORDS

Therapeutic adherence; therapeutic competence; treatment integrity; prolonged grief disorder; CBT

PALABRAS CLAVES

Adherencia terapéutica; competencia terapéutica; integridad de tratamiento; trastorno por duelo prolongado; TCC

关键词

治疗依从性; 治疗能力; 治疗完整性; 延长哀伤障碍; CBT

HIGHLIGHTS

- This is the first study to report on the development and psychometrical evaluation of a grief-focused adherence and competence scale.
- Although an increasing number of clinical trials do report the efficacy of grief-focused cognitive-behavioural therapy, none of these studies used a standardized adherence and competence scale to control internal validity.
- In the present study, we introduced a therapeutic adherence and competence scale for grief (TACs-G) that can be applied efficiently across different research settings (e.g. manipulation check, dissemination), and report results of good to excellent psychometric properties.
- The scale itself could prove useful beyond the research setting as it could possibly serve as a basis for feedback in training settings.

CONTACT Julia Haneveld ✉ julia.haneveld@ku.de Department of Psychology, Catholic University Eichstaett-Ingolstadt, Levelingstr. 7, Ingolstadt 85049, Germany

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/20008198.2022.2079873>.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

individuales (ICC = 0.509 a 1.00). Se encontró que la estabilidad de los TACs-G durante el tiempo fue fuerte tanto para los puntajes totales de adherencia (ICC = 0.970) como de competencia (ICC = 0.965).

Conclusiones: La TAC-G para TCC es un instrumento confiable que puede ser usado no solamente para asegurar validez interna sino también para estudios sobre proceso psicoterapéutico. Además, provee de una base de datos valiosa para retroalimentación dirigida en contextos de capacitación.

聚焦哀伤的认知行为治疗的治疗依从性和能力量表的介绍和评估

背景: 没有聚焦哀伤的认知行为疗法 (聚焦哀伤CBT) 的治疗能力和依从性量表。然而, 鉴于越来越多的证据表明聚焦哀伤 CBT 的效度, 需要这样的量表来确保临床试验的内部效度和促进心理治疗过程的研究。

目的: 开发和开展一项针对聚焦哀伤 CBT 的治疗依从性和能力量表的心理测量评估。

方法: 该量表分两步开发。(I) 五位延长哀伤障碍治疗专家对条目的相关性和适当性进行了反馈。对量表进行了修订以反映他们的反馈。最终的哀伤治疗依从性和能力量表 (TACs-G) 由 15 个依从性和 16 个能力条目组成。(II) TACs-G 的心理测量评估基于两名独立评分者对 48 个随机选择的 PG-CBT 会话的评分。这些视频是在随机对照试验 (RCT; DRKS00012317) 的背景下录制的。ICC 用于计算评分者间的可靠性和 TACs-G 随时间的稳定性 (12 个月重新评分 10 个会话)。

结果: 五位专家确认了条目的相关性和适当性。总体依从性和能力得分 (ICC = 0.889 和 0.782) 的评分者间信度较高, 个别条目的信度为中等至极好 (ICC = 0.509 至 1.00)。随着时间的推移, TACs-G 的稳定性对于依从性 (ICC = 0.970) 和能力总分 (ICC = 0.965) 都很强。

结论: 用于 CBT 的 TACs-G 是一种可靠的工具, 不仅可用于确保内部效度, 还适用于心理治疗过程研究。此外, 它还还为培训设置中的针对性反馈提供了一个有价值的数据库。

1. Introduction

By March 2022, there had been more than 6.1 million COVID deaths worldwide (WHO, 2022). Given that each deceased person leaves behind at least nine relatives (Verdery, Smith-Greenaway, Margolis, & Daw, 2020) and that these deaths mostly occurred suddenly, unexpectedly, and with little or no time for the bereaved to say goodbye, it is not surprising that an increase in prolonged grief disorder (PGD) can be assumed (Eisma & Boelen, 2021). Accordingly, the need for effective and grief-specific treatments has intensified.

Over the past decade, PGD has been extensively investigated and has recently been included in ICD-11 and DSM-5-TR in 2018 and 2022, respectively (APA, 2022; WHO, 2018). In parallel to the debate about diagnostic criteria for this disorder, considerable efforts have been devoted to researching effective therapeutic approaches. In a recent meta-analysis Johannsen et al. (2019) reported medium effect sizes ($g = 0.41$) for grief-specific psychological interventions (including Cognitive Behavioural Therapy (CBT) but also emotional supportive and psychoeducational approaches). Randomized controlled trials (RCTs) on the efficacy of CBT interventions for PGD even observed large effects (Boelen, de Keijser, van den Hout, & van den Bout, 2007; Bryant et al., 2014; Bryant et al., 2017; Rosner, Bartl, Pfoh, Kotoučová, & Hagl, 2015; Rosner, Lumbeck, & Geissner, 2011; Rosner, Pfoh, Kotoučová, & Hagl, 2014; Shear et al., 2014; Shear et al., 2016). However, none of these studies used state-of-the-art techniques to control whether the intervention was implemented as planned. In

scientific research, such a procedure is broadly referred to as controlling for treatment integrity. According to Barber, Triffleman, & Marmar (2007), state-of-the-art techniques involve the rating of audiotapes or videotapes using psychometrically evaluated scales by independent raters.

This lack of verification is not uncommon in psychotherapy research. In their review, Perepletchikova, Treat, and Kazdin (2007) reported that treatment integrity was adequately addressed in only 3.5% of trials on psychosocial interventions published in the most influential psychiatric and psychological journals. Nevertheless, a high level of treatment integrity is a prerequisite for ensuring the internal validity of psychotherapy studies (Perepletchikova & Kazdin, 2005; Schlosser, 2002; Weck, Grikscheit, Höfling, & Stangier, 2014).

Treatment integrity encompasses treatment differentiation, adherence, and competence (Waltz, Addis, Koerner, & Jacobson, 1993). Treatment differentiation requires that different treatments are distinguishable from one another in terms of critical aspects (e.g. the therapist of a depth psychology-oriented therapy refrains from behavioural expositions; Weck, Bohn, Ginzburg, & Stangier, 2011). Competence pertains to the skill level demonstrated by the therapist when delivering the treatment as well as how the therapist interacts with the patient (Weck et al., 2011). Adherence describes the extent to which a therapist uses interventions and approaches prescribed by the treatment manual and avoids the use of intervention procedures proscribed by the manual (Waltz et al., 1993). Thus, the adherence measurement ensures internal validity in clinical trials (Moncher & Prinz, 1991; Waltz

et al., 1993) as changes in the dependent variable (treatment outcome) can be attributed to the independent variable (intervention). In addition, external validity is achieved when results are generalizable. This is not the case when the independent variable (intervention) is not replicable because the intervention was not carried out as planned (Weck et al., 2014). Thus, adherence ratings are recommended in textbooks on psychotherapy research (Lambert, 2013). However, adherence should not only be addressed as a methodological requirement. It could also be a predictor (assuming that the elements prescribed in the manual are the active components in therapy, adherence should positively predict outcome), moderator (higher adherence could lead to more improvement regarding the outcome), or mediator (adherence might establish or explain the relationship between intervention and outcome). Thus, the study of adherence can be helpful for the interpretation of outcomes, and can contribute to the accurate identification of therapy-specific, active components (e.g. exposure or cognitive restructuring; Barber, Sharpless, Klostermann, & McCarthy, 2007). Finally, a well-designed and evaluated grief-focused adherence and competence scale could be used beyond the research setting in therapists' skills training.

To the best of our knowledge, no adherence or competence scale has been developed for any grief-focused preventive or therapeutic intervention so far. Thus, with the aim of advancing the field of PGD treatment research, we developed a grief-focused adherence and competence scale for CBT in this study. The basis for our scale structure was the adherence and competence scale for developmentally adapted cognitive processing therapy for adolescents with post-traumatic stress disorder (PTSD) developed by Gutermann et al. (2015). The content of the scale was based on the integrative CBT for prolonged grief disorder (PG-CBT) manual by Rosner, Pfoh, Kotoučová, and Comtesse (2015). PG-CBT can be considered a proxy for other grief-focused CBT approaches, and has been shown to be effective in a pilot trial comprising 51 bereaved adults with PGD. Compared with a waiting list, between group effect sizes were large for the improvement of grief symptoms in the intent-to-treat analysis ($d = 1.32$; Rosner et al., 2014).

The primary goal of our study was the development and psychometric evaluation of a therapeutic adherence and competence scale for grief-focused CBT, TACs-G. We present the two steps of the TACs-G development, including expert ratings, to ensure content validity (first step) and psychometric evaluation by examining its interrater reliability and its stability over time (second step). The TACs-G psychometric evaluation was based on the rating of 48 PG-CBT videotaped sessions by two independent raters. Finally, we discuss whether the scale could be

employed as an efficient rating instrument based on the average rating time, and outline how it provides a database for process research purposes.

2. Development of the TACs-G

As there are no grief-focused CBT adherence and competence scales, we screened the literature for adherence and competence scales for other disorders. Among others, we found scales for depression, eating disorders, anxiety disorders, and post-traumatic stress disorder (PTSD; Folke et al., 2017; Ginzburg et al., 2012; Gutermann et al., 2015; Shaw et al., 1999; von Consbruch, Clark, & Stangier, 2012). Due to similarities in symptoms between PGD and PTSD (e.g. avoidance of reminders; see Djelantik, Smid, Kleber, & Boelen, 2017), we decided to use the scale of Gutermann et al. (2015) as a structural basis for our scale.

The PG-CBT by Rosner, Pfoh, Rojas, et al. (2015) served as a content basis, as it comprises a range of different session-specific CBT interventions (e.g. introduction of a personalized disturbance model, cognitive restructuring). It, therefore, complies with the request of Perepletchikova et al. (2007) to measure treatment integrity based on a manual with session-specific descriptions. In addition to this, PG-CBT shares certain similarities with other grief-focused manuals (e.g. in sensu exposure, cognitive restructuring of dysfunctional grief-related thoughts, see for example, Shear, 2015). Consequently, the scale is applicable beyond the context of testing adherence to the PG-CBT manual. In the following, we provide detailed information on the structural and content basis of the scale, and we describe the item validation based on an expert survey (step I of the scale development).

2.1. Content basis of the scale

The scale's content was informed by the PG-CBT manual by Rosner, Pfoh, Rojas, et al. (2015). PG-CBT comprises 24 individual, weekly sessions (see session content description in Supplement Table A). 20 of these 24 sessions are divided into three phases (A, B, C). The remaining four sessions are optional sessions, which are reserved for discussing specific events (e.g. court hearing for inheritance disputes or anniversaries). Each phase entails a different focus in grief treatment. Phase A comprises stabilizing, exploring, motivating, and goal setting. It starts with an open clinical talk, which focuses on coping mechanisms employed by the patient to deal with their grief. Furthermore, the patient's grief behaviour is identified by means of a disturbance model. In this context, family experiences of dealing with grief (learned grief behaviour) and secondary losses are also revealed. In addition, the deceased is introduced,

whereby one-sided representations (e.g. idealization) are questioned. At the end of phase A, therapy goals are formulated and the motivation to change is encouraged.

Phase B concentrates on re-interpreting and confronting grief. It is, therefore, the central aspect of the PG-CBT. As part of cognitive restructuring, patients are asked to name their most stressful thoughts. Various techniques are used for cognitive restructuring. Furthermore, the exposure takes place in this phase. During exposure the patient is confronted with the worst moment regarding the loss. Finally, phase B finishes with the 'visiting the grave' exercise. The goal of this exercise is reconciliation or correction of unfinished situations and obtaining permission to continue with one's life.

Phase C deals with integrating and transforming grief into concluding thoughts. After the patient has received the deceased's permission to create a future life with a new goal orientation (during the 'visiting the grave' exercise), the focus shifts to the patient's future roles, hopes, and wishes. In addition, the patient decides what he wants to keep from the deceased, and what he wants to leave behind (material as well as immaterial). The last sessions revolve around the end of the treatment; accordingly, achievements are reflected upon, and relapse prevention is carried out.

2.2. Structural basis of the scale

The Adherence (TAS) and Competence Scale (TCS) by Gutermann et al. (2015) served as the inspiration for the scale's structure and the wording of the items. The TAS and TCS are based on a cognitive processing therapy manual for adolescents suffering from PTSD. The TCS consists of 21 items, which are derived from the Cognitive Therapy Scale by Weck, Hautzinger, Heidenreich, and Stangier (2010), as well as 7 treatment-specific competence items. The TAS comprises 12 items, which can be rated in each session irrespective of the therapy phase. Hence, in order to include phase-specific elements (e.g. reading and editing the trauma report) which are not the content of each session, Gutermann et al. (2015) had to create one item that summarizes the 44 phase-specific interventions. Consequently, the advantage of an uncomplicated rating comes at the cost of information loss. In this respect, our scale differed significantly from that of Gutermann et al. (2015), as we wanted to capture phase-specific interventions in detail. The correct and detailed recording of phase-specific intervention was critical in this study because the scale is to be used not only to ensure internal validity of the treatment, but also to create a database for process research. In this regard, we also deemed it important to clearly differentiate between adherence and

competence. As Barber, Triffleman, et al. (2007) and Dobson and Singer (2005) have pointed out, these two constructs have often been mistaken for each other, making it difficult to clearly attribute the outcome effects to adherence or competence, or to identify interaction effects. While adherence ratings examine the extent to which a treatment was carried out, competence ratings address the manner in which it was carried out (Barber, Triffleman, et al., 2007). Establishing a sound rapport with the patient, for example, is more a matter of competence than adherence, yet this aspect is often assessed as part of adherence measurement (see for example Dittmann et al., 2017). To avoid such interference, we followed recommendations from previous work (Barber, Triffleman, et al., 2007; Dobson & Singer, 2005) to measure adherence separately from competence. Consequently, the scale comprised adherence items with separate but associated competence items.

2.3. Item validation

For content validation, we asked five independent German experts involved in the treatment of PGD to provide feedback on the items in the TACs-G. The experts had on average 17 years' (SD = 7.9) clinical experience. The expert survey itself was based on the Haynes, Richard, and Kubany (1995) definition and recommendation for content validation. Thus, the experts were asked to evaluate the relevance and appropriateness of each item on a scale of 1 (not at all relevant/appropriate) to 5 (extremely relevant/appropriate). Specifically, the experts were asked to estimate whether the items represented relevant content for grief treatment and whether they were suited to capturing the therapist's adherence/competence for the content in question. Furthermore, we asked the experts to freely comment on each item so as to generate more holistic feedback. Their feedback helped us to shorten the scale and to emphasize the difference between the adherence and competence items. In this regard, two items were formulated as pure competence items, namely, to deal with avoidance behaviour and grief-specific resource activation. The experts considered these two aspects to be a matter of skill rather than of adherence. After integrating the PGD experts' feedback, we initiated a focus group discussion and asked a group of experts in the field of psychotraumatology and psychotherapy research for feedback. Their feedback not only improved the clarity and completeness of the items, but also eliminated redundancies between items to ensure item independence (Table D in the Supplement provides details of the modification which was undertaken based on the experts' feedback). This intermediate step was useful in the context of our intention to develop a resource-saving scale that could be easily understood

by less trained raters. Finally, in a last step, we contacted our initial experts again, presented our final scale and asked them to rate the relevance and appropriateness of each item (see [Tables 1 and 2](#)).

2.4. The final TACs-G

Overall, the TACs-G comprised 15 adherence and 16 competence items. For 14 adherence items, we developed matching competence items so as to clearly distinguish between adherence and competence. The 15th adherence item comprised a list of proscribed interventions, so no associated competence item was needed. Two additional competence items, namely 'grief specific resource activation' and 'dealing with avoidance behaviour' were included. For the adherence items a three-point Likert scale was used (1 = intervention is not implemented although indicated, 2 = intervention is partially implemented, 3 = intervention is implemented). For the competence scale a five-point Likert scale was used, ranging from 1 = completely insufficient competence to 5 = very good (see Supplement Table C for an overview of the TACs-G items). In addition to the TACs-G itself, an extensive rater manual with various explanations and examples was developed (Haneveld & Comtesse, 2020). During the rating process, the two raters could always refer to it in the event of any uncertainties.

3. Methods regarding psychometric evaluation of the TACs-G

3.1. Treatment

The data for this study were obtained from an ongoing multicenter RCT to evaluate grief-focused CBT in comparison with an active control condition, Present-Centered Therapy (PCT; German Clinical Trials Register, ID: DRKS00012317). The study protocol has been approved by the IRB of the Catholic University Eichstaett- Ingolstadt (2016/21). Eligible participants had to meet the criteria of a primary PGD diagnosis according to Prigerson et al. (2009) criteria as assessed using the Prolonged Grief-13 + 9 interview (Vogel, Pfoh, & Rosner, 2017). Moreover, to qualify for inclusion in the trial, participants had to be between 18 and 75 years old, and their loss had to have occurred at least 6 months before (for all inclusion criteria see the study protocol: Rosner, Rimane, Vogel, Rau, & Hagl, 2018).

3.2. Video-taped patients and therapists

This study used the data of patients who were included as pilot cases ($n = 8$; i.e. each therapist underwent intensive training and had to treat a supervised pilot

case before entering the main trial) and data of main trial patients ($n = 8$). Most of the patients were female (69%). The average age was 51.2 years ($SD = 9.3$; range 34-69). Regarding psychiatric comorbidities, the majority (81.25%) had a co-morbid mental disorder, typically major depression or anxiety disorders according to the SKID-I interview (Wittchen, Zaudig, & Fydrich, 1997; for more details refer to Supplement Table B).

Most of the therapists were female (94%) and had on average 39.7 months' clinical experience ($SD = 19.42$). All of them were clinical psychologists (12 of them were undergoing training to become a licensed therapist and 3 were licensed therapists). They had been trained in PG-CBT by the developers of the treatment manual in a 2-day workshop, and were supervised at least once every four weeks.

3.3. Raters and sampling

The ratings were undertaken by two CBT therapists (the first and fourth author of this paper). The first was a licensed psychotherapist and the second was undergoing training in psychotherapy (4 and 2 years' clinical experience, respectively). Both raters had received training on the PG-CBT manual in a 2-day workshop (see Rosner et al., 2018). They had also attended an additional 4-hour rater training, based on the rater manual for the TACs-G (Haneveld & Comtesse, 2020). Additionally, the raters had to evaluate 10 practice videos and to reach a consensus before starting to rate the videos included in this study. To minimize rater drift, the two raters met throughout the study to discuss the videos.

The sampling was done in two steps. In a first step, we randomly selected 16 treatments (33%) from the total sample of 48 videotaped treatments (each treatment consisted of 20 sessions; the optional sessions, a maximum of four, were excluded from the sampling because their content was based on the individual patient's need and thus not prescribed in the manual). In a second step, we randomly selected one session per phase (A, B, C respectively), from each treatment. Thus, our sample resulted in 3 sessions (representing phase A, B, C) from 16 treatments (48 sessions in total). The 16 treatments, in turn, were provided by 15 therapists. The therapists did not know which phase or which session would be selected for rating. If a video from a particular session was missing or damaged, an adjacent session was used (please refer to the Supplement, Figure A for an illustration of the stratified sampling procedure).

3.4. Statistical analyses

To determine the reliability of the TACs-G scale, we calculated both interrater and intra-rater reliability.

The latter should account for the scale's stability over time. Accordingly, re-evaluation of 10 sessions took place after 12 months – to minimize memory effects – and was conducted by the first author. To evaluate the efficiency of the TAC-G scale, the average rating time of the initial rating was analyzed. All data were analyzed using RStudio, version 1.2.5042 (RStudio, 2020).

Both inter- and intra-rater reliability were determined by calculating the intraclass correlation coefficient (ICC) using Model 2 [ICC_(2,1)], according to Shrout and Fleiss (1979). The 95% confidence interval was used to determine statistical significance. According to Portney and Watkins (2009), ICCs > 0.80 can be categorized as excellent, ICCs between 0.70 and 0.80 as good, ICCs between 0.50 and 0.60 as moderate and <0.5 as not satisfactory. The ICC for the total adherence/total competence score was calculated based on the total sum score of items 1–14/1–16 per rating. Pearson's product moment correlation (*r*) was used to determine the correlation between the adherence and the competence scales.

4. Results

Tables 1 and 2 display the results of the experts' ratings for each item as well as the ICCs, the means, and the range. The five experts considered all the *adherence* items as appropriate with $M = 4.60$ ($SD = 0.21$; range 1–5) and relevant with $M = 4.81$ ($SD = 0.26$; range 1–5). Likewise, they deemed all the *competence* items to be appropriate with $M = 4.60$ ($SD = 0.22$; range 1–5) and relevant with $M = 4.81$ ($SD = 0.25$; range 1–5). Furthermore, they considered all phase-specific adherence and most phase-specific competency items ($M = 5.0$) to be particularly relevant.

The interrater reliability for the total adherence and total competence scores was good to excellent (for adherence: ICC_{2,1} = 0.889, for competence: ICC_{2,1} = 0.782) according to Portney and Watkins (2009). With respect to the individual adherence items, the raters showed excellent agreement with one exception (Adherence Item 1, Agenda: ICC_{2,1} = 0.796). Regarding the individual competence items, all but two items achieved excellent interrater reliability. Only competence items 15 (avoidance behaviour) and 16 (grief-specific resource activation), evidenced moderate agreement (Item 15: ICC_{2,1} = 0.509 and Item 16: ICC_{2,1} = 0.678). Furthermore, there was a strong positive correlation between the sum scores of the adherence and competence scales, $r = 0.80$ $p < 0.001$.

To test for the scale's stability over time, 10 videos were randomly drawn from the sample and were re-scored by the first author. The average time between ratings was 13.1 months (range 13–14 months). Results yielded high intra-rater reliability and thus high stability over time for the sum scores of the

adherence (ICC_{2,1} = 0.970) and the competence scales (ICC_{2,1} = 0.965). As can be seen in Tables 3 and 4, the intra-rater reliability for the individual adherence and competence items ranged from moderate to excellent (from 0.724 to 1.0, and from 0.530 to 1.0, respectively).

To assess the TACs-G efficiency, the rating time was analyzed. On average, it took 13.9 additional minutes ($SD = 4.8$, range from 5 to 25 min) after having watched the therapy video to finish the TACs-G rating.

5. Discussion

The aim of this study was to develop and evaluate an adherence and competence scale for grief-focused CBT. Overall, the results indicated that the TACs-G was appropriate, relevant, and efficient. It allowed a reliable assessment of therapeutic adherence and competence.

As this is the first adherence and competence scale for grief-focused CBT, it was important for us to include independent experts in this study, and to collect their feedback on item content. Overall, the five experts considered all adherence and competence items to be relevant and appropriate. More particularly, they deemed phase-specific adherence and most phase-specific competence items to be highly relevant. This ultimately reinforced our initial decision to include phase-specific elements separately rather than grouping them together in one item. Our scale was efficient, as ratings were completed relatively fast (average completion time of 14 min for 31 items) compared to previous work. Dittmann et al. (2017), for example, reported a completion time of 10 min for a pure adherence scale with 11 items. The rating time may still vary depending on the experience of the individual rater, which was not examined in this study.

Our study observed good to excellent rater accordance for the total TACs-G adherence and competence scores as well as for the individual items. This result compared favourably with the high rater ICCs found for other integrity measures of CBT for Anxiety Disorders or PTSD (see Barber, Liese, & Abrams, 2003; Bjaastad et al., 2016; Gutermann et al., 2015). Three adherence items (nos. 9, 11, 14) and one competence item (no. 9) showed a particularly high ICC coefficient of 1.00. The overall high ICC coefficients could perhaps be explained by the low to no level of ambiguity in the items, and/or by the clear instructions in the rater manual. Another factor might be the use of only two (highly trained) raters. Yet, many studies that evaluated adherence scales reported on data based on two raters (see Bjaastad et al., 2016; Folke et al., 2017; Gutermann et al., 2015). Finally, the high ICCs of the adherence items could also be due to the use of restricted scales (ranging from 1 to 3), leaving only a few options with little nuance.

Table 1. Therapeutic Adherence and Competence scale for Grief-focused cognitive behavioural therapy (TACs-G): *Adherence items only*: intraclass correlation coefficient, number of ratings, mean and standard deviation of items, range and mean and standard deviation of final expert ratings.

| Item | ICC _{2,1} (CI 95%) | N | Mean (SD) | Min/Max | Appropriateness Mean (SD) | Relevance Mean (SD) |
|--|-----------------------------|----|--------------|---------|---------------------------|---------------------|
| 1. Agenda | .796* (.663–.880) | 96 | 2.51 (0.68) | 1/3 | 4.20 (0.45) | 4.60 (0.55) |
| 2. Time management | .925* (.870–.957) | 96 | 2.22 (0.84) | 1/3 | 4.20 (0.84) | 4.60 (0.55) |
| 3. Dealing with ambivalence | .800* (.670–.883) | 96 | 2.68 (0.61) | 1/3 | 4.20 (0.84) | 4.80 (0.45) |
| 4. Dealing with/addressing stressors | .989* (.980–.994) | 32 | 2.75 (0.44) | 2/3 | 4.20 (0.45) | 4.40 (0.55) |
| 5. Psychoeducation regarding PGD, comorbidities and PG-CBT | .994* (.990–.997) | 34 | 2.74 (0.45) | 2/3 | 4.80 (0.45) | 5.00 (0.00) |
| 6. Working with cognitions | .907* (.840–.947) | 42 | 2.57 (0.55) | 1/3 | 4.80 (0.45) | 4.80 (0.45) |
| 7. Feelings of grief/Perception of emotions | .938* (.893–.965) | 23 | 2.83 (0.39) | 2/3 | 4.60 (0.55) | 4.60 (0.55) |
| 8. Promoting self-help | .822* (.703–.896) | 51 | 2.51 (0.70) | 1/3 | 4.80 (0.45) | 5.00 (0.00) |
| 9.A Getting to know the patient and his grief reaction | 1.0* (1.0–1.0) | 2 | 3.00 (0.00) | 3/3 | 4.40 (0.55) | 5.00 (0.00) |
| 10.A. Introducing the deceased person | .953* (.918–.973) | 11 | 2.91 (0.30) | 2/3 | 5.00 (0.00) | 5.00 (0.00) |
| 11.B Exposure: Worst moment | 1.0 (1.0–1.0) | 12 | 2.83 (0.39) | 2/3 | 5.00 (0.00) | 5.00 (0.00) |
| 12.B Visiting the grave | .805* (.677–.886) | 14 | 2.79 (0.43) | 2/3 | 4.60 (0.55) | 5.00 (0.00) |
| 13.C Legacy, remembrance, dedication | .870* (.780–.925) | 18 | 2.67 (0.49) | 2/3 | 4.80 (0.45) | 5.00 (0.00) |
| 14.C Future prospects | 1.0 (1.0–1.0) | 18 | 2.67 (0.49) | 2/3 | 4.80 (0.45) | 5.00 (0.00) |
| 15. Proscribed interventions | 1.0 (1.0–1.0) | 96 | 1.00 (0.00) | 1/1 | 4.25 (0.50) | 4.00 (0.82) |
| Total score | .889* (.810–.936) | 96 | 14.58 (3.11) | 7/23 | | |

The letter (A, B, C) next to the item number refers to the therapy phase, ICC_{2,1} = intraclass correlation coefficients for both raters, Min = lowest rating score on a scale of 1–3, Max = highest rating score on a scale of 1–3. Relevance and appropriateness were assessed on a scale ranging from 1 to 5. * $p < .001$.

In contrast to these items with very high ICC coefficients, there were also two competence items (nos. 15 and 16) that only showed moderate interrater reliability. One possible explanation for this might be that these two items were the only ones formulated as pure competence items without any associated adherence items. Thus, the raters were forced to first consider whether the intervention followed the manual and then the extent *to which* it was implemented. Following this, the competence item had to be rated in terms of *how* this intervention was implemented (completely insufficient to very good). Hence, the lack of guidance, which was otherwise provided by the adherence item, may have made the rating more difficult. Finally, other studies reported similar moderate ICC coefficients for competence scales which were not associated with an adherence rating (refer to Bjaastad et al., 2016; von Consbruch et al., 2012). This suggested that reliably estimating competence was more demanding than estimating adherence, and was also evinced by higher overall interrater reliability of adherence compared to competence (ICC = 0.889 vs. ICC = 0.782). Finally, in this context the raters' 'limited' clinical experience may have had an impact on their ability to recognize varying skill levels, and thus to assess the therapists' competence. However, other studies also employed raters with 'limited' experience by using the master students (see von Consbruch et al., 2012; Gutermann et al., 2015; Ginzburg et al., 2012; Loeb et al., 2005; Meier et al., 2015; Weck et al., 2014) and our raters were, after all, were a psychotherapist and a psychotherapist in training.

An internal consistency analysis was deliberately not undertaken in this study for two reasons. First, this would have constituted a violation of the prerequisite for such an analysis, meaning no underlying unidimensional construct could be assumed due to

the phase-specific and thus heterogenous structure (Schermelleh-Engel & Werner, 2012). Second, in other studies, the utility and, by extension, the reasonable interpretation of Cronbach's alpha in the context of treatment integrity scales has been questioned. The recommendation given was to use alternative reliability measures whenever possible (refer to the discussion of Dittmann et al., 2017; Gutermann et al., 2015; von Consbruch et al., 2012). Hence, we employed ICC_{2,1} to determine intra-rater reliability and thus the stability of the scale over time. Results yielded a strong 12-month intra-rater reliability for the TACs-G adherence, competence scores and individual items. A favourable interpretation of this result is that the scale contains precisely formulated items that allow accurate and consistent measurement of adherence and competence over time. However, memory effects cannot be ruled out. To reduce possible impacts of memory effects, future studies should consider a longer time period in between the ratings.

We observed a very strong positive correlation between TACs-G adherence and competence scores. There are three possible reasons for this. First, therapists who are good at adhering to the treatment protocol are skilled therapists with a high level of competence (Bjaastad et al., 2016). Second, raters could possibly have difficulties separating adherence from competence in their ratings (Bjaastad et al., 2016). However, in our case, raters underwent intensive training in which the difference was emphasized. In addition, the rater manual devoted an entire section to the difference between competence and adherence, which is expected to minimize rater uncertainty in this study. Previous work (Barber et al., 2003; Bjaastad et al., 2016) also found a very strong association ($r = 0.96$ and $r = 0.70$, respectively). In both studies, the differences between adherence and competence

Table 2. Therapeutic Adherence and Competence scale for Grief-focused cognitive-behavioural therapy (TACs-G): *Competence items only*: intraclass correlation coefficient, number of ratings, mean and standard deviation of items, range and mean and standard deviation of final expert ratings.

| Item | ICC _{2,1} (CI 95%) | N | Mean (SD) | Min/Max | Appropriateness Mean (SD) | Relevance Mean (SD) |
|--|-----------------------------|----|--------------|---------|---------------------------|---------------------|
| 1. Agenda | .946* (.906–.969) | 96 | 3.58 (1.07) | 1/5 | 4.60 (0.55) | 4.60 (0.55) |
| 2. Time management | .916* (.854–.952) | 96 | 3.83 (0.84) | 2/5 | 4.40 (0.55) | 4.80 (0.45) |
| 3. Dealing with ambivalence | .901* (.831–.943) | 96 | 3.76 (1.10) | 1/5 | 4.40 (0.89) | 4.80 (0.45) |
| 4. Dealing with/addressing stressors | .988* (.979–.993) | 32 | 4.41 (0.67) | 3/5 | 4.60 (0.55) | 4.60 (0.55) |
| 5. Psychoeducation regarding PGD, comorbidities and PG-CBT | .988* (.979–.993) | 34 | 4.21 (0.81) | 3/5 | 4.80 (0.45) | 5.00 (0.00) |
| 6. Working with cognitions | .903* (.834–.945) | 42 | 3.93 (0.97) | 2/5 | 4.80 (0.45) | 4.80 (0.45) |
| 7. Feelings of grief/Perception of emotions | .918* (.858–.953) | 23 | 4.26 (0.92) | 2/5 | 4.00 (0.71) | 4.60 (0.55) |
| 8. Promoting self-help | .825* (.707–.898) | 51 | 3.39 (1.19) | 1/5 | 4.60 (0.55) | 4.80 (0.45) |
| 9.A Getting to know the patient and his grief reaction | 1.0* (1.0–1.0) | 2 | 4.00 (0.00) | 4/4 | 4.60 (0.55) | 4.80 (0.45) |
| 10.A Introducing the deceased | .892* (.816–.938) | 11 | 4.82 (0.41) | 4/5 | 5.00 (0.00) | 5.00 (0.00) |
| 11.B Exposure: Worst moment | .969* (.946–.983) | 12 | 4.17 (1.03) | 2/5 | 4.60 (0.55) | 5.00 (0.00) |
| 12.B Visiting the grave | .824* (.707–.898) | 14 | 3.86 (0.95) | 2/5 | 4.80 (0.45) | 4.80 (0.45) |
| 13.C Legacy, remembrance, dedication | .815* (.692–.892) | 18 | 4.22 (0.88) | 2/5 | 4.80 (0.45) | 5.00 (0.00) |
| 14.C Future prospects | .951* (.914–.972) | 18 | 4.39 (0.85) | 2/5 | 4.20 (0.45) | 4.60 (0.55) |
| 15. Dealing with 'avoidance behaviour' | .509* (.266–.692) | 57 | 3.75 (1.14) | 2/5 | 5.00 (0.00) | 5.00 (0.00) |
| 16. Grief specific resource activation | .678* (.490–.806) | 70 | 3.67 (1.05) | 1/5 | 4.40 (0.55) | 4.80 (0.45) |
| Total score | .782* (.642–.872) | 96 | 26.92 (6.61) | 13/45 | | |

The letter (A, B, C) next to the item number refers to the therapy phase, ICC_{2,1} = intraclass correlation coefficients for both raters, Min = lowest rating score on a scale of 1–5, Max = highest rating score on a scale of 1–5. Relevance and appropriateness were assessed on a scale ranging from 1 to 5. **p* < .001.

were stressed in the raters' training as well. Altogether, these results suggested that rater difficulty in separating adherence from competence was not a determining factor in the correlation between adherence and competence scores. Finally, the same raters assessed both adherence and competence. Having different raters assess adherence and competence would shed light on whether the high correlations were an artefact of using the same rater.

Besides the constraints already discussed, further limitations should be kept in mind when interpreting

Table 3. Stability over time for the Therapeutic Adherence and Competence scale for Grief-focused cognitive behavioural therapy (TACs-G), *Adherence items only*.

| Item | ICC _{2,1} (CI 95%) | N | Mean (SD) | Min/Max |
|--|-----------------------------|----|--------------|---------|
| 1. Agenda | 1.000 | 20 | 2.70 (0.47) | 2/3 |
| 2. Time management | 1.000 | 20 | 2.30 (0.92) | 1/3 |
| 3. Dealing with ambivalence | 0.724 | 20 | 2.35 (0.75) | 1/3 |
| 4. Dealing with/addressing stressors | 0.974 | 6 | 2.82 (0.41) | 2/3 |
| 5. Psychoeducation regarding PGD, comorbidities/& PG-CBT | 1.000 | 6 | 3.00 (0.00) | 3/3 |
| 6. Working with cognitions | 1.000 | 8 | 2.00 (0.76) | 1/3 |
| 7. Feelings of grief/ Perception of emotions | 1.000 | 2 | 2.00 (0.00) | 2/2 |
| 8. Promoting self-help | 0.972 | 11 | 2.09 (1.04) | 1/3 |
| 9.A Getting to know the patient and his grief reaction | – | 0 | – | – |
| 10.A Introducing the deceased | 1.000 | 4 | 3.00 (0.00) | 3/3 |
| 11.B Exposure: Worst moment | 1.000 | 4 | 2.50 (0.57) | 2/3 |
| 12.B Visiting the grave | 0.947 | 4 | 2.25 (0.50) | 2/3 |
| 13.C Legacy, remembrance, dedication | 0.923 | 2 | 2.50 (0.71) | 2/3 |
| 14.C Future prospects | 0.964 | 4 | 2.75 (0.50) | 2/3 |
| 15. Proscribed interventions | 1.000 | 20 | 1.00 (0.00) | 1/1 |
| Total score | 0.970 | 20 | 13.60 (3.73) | 7/21 |

The letter (A, B, C) next to the item number refers to the therapy phase, ICC_{2,1} = Intraclass correlation coefficients Model 2, Min = lowest rating score on a scale of 1–3, Max = highest rating score on a scale of 1–3.

the results of this study. The major limitation was the relatively small sample size with 48 video ratings in total. It should, nonetheless, be noted that this sample size was already larger than in comparable scale development studies (e.g. Dittmann et al., 2017; Gutermann et al., 2015 with *N* = 30). A recent meta-analysis by Zarafonitis-Müller, Kuhr, and Bechdorf (2014) focused on CBT studies investigating adherence and

Table 4. Retest Reliability for the Therapeutic Adherence and Competence scale for Grief-focused cognitive behavioural therapy (TACs-G), *Competence items only*.

| Item | ICC _{2,1} (CI 95%) | N | Mean | Min/Max |
|--|-----------------------------|----|--------------|---------|
| 1. Agenda | 0.685 | 20 | 4.00 (0.97) | 2/5 |
| 2. Time management | 0.835 | 20 | 4.00 (1.08) | 2/5 |
| 3. Dealing with ambivalence | 0.925 | 20 | 3.05 (1.43) | 1/5 |
| 4. Dealing with/addressing stressors | 1.000 | 6 | 4.33 (0.52) | 4/5 |
| 5. Psychoeducation regarding PGD, comorbidities & PG-CBT | 0.991 | 6 | 4.83 (0.41) | 4/5 |
| 6. Working with cognitions | 0.882 | 8 | 2.75 (1.28) | 1/5 |
| 7. Feelings of grief/ Perception of emotions | 1.000 | 2 | 3.00 (0.00) | 3/3 |
| 8. Promoting self-help | 0.951 | 11 | 2.81 (1.40) | 1/5 |
| 9.A Getting to know the patient and his grief reaction | – | 0 | – | – |
| 10.A Introducing the deceased | 0.985 | 4 | 4.25 (0.50) | 4/5 |
| 11.B Exposure: Worst moment | 0.983 | 4 | 3.75 (1.50) | 2/5 |
| 12.B Visiting the grave | 0.948 | 4 | 3.00 (0.82) | 2/4 |
| 13.C Legacy, remembrance, dedication | 1.000 | 2 | 4.00 (0.00) | 4/4 |
| 14.C Future prospects | 1.000 | 4 | 4.00 (1.55) | 3/5 |
| 15. Dealing with 'avoidance behaviour' | 0.530 | 14 | 3.07 (1.28) | 2/5 |
| 16. Grief-specific resource activation | 0.985 | 14 | 3.50 (0.85) | 2/5 |
| Total score | 0.965 | 20 | 24.65 (8.01) | 16/44 |

The letter (A, B, C) next to the item number refers to the therapy phase, ICC_{2,1} = Intraclass correlation coefficients Model 2, Min = lowest rating score on a scale of 1–3; Max = highest rating score on a scale of 1–3.

competence based on audiotape or videotape ratings, and reported sample sizes ranging from 25 to 62. Thus, our sample size was consistent with previous work. In general, the small number of integrity studies and the relatively small sample sizes reported in these studies were attributable to the extensive resources required (financial and time commitments) to develop and apply adherence and competence measures (Webb, Derubeis, & Barber, 2010; Zarafonitis-Müller et al., 2014).

Another limitation was the relative homogeneity of the videotaped therapists and patients (e.g. mostly female). Consequently, the results should be confirmed using a larger and more heterogeneous sample of therapists and patients. Of course, this aspect is of even greater importance when conducting a process research study. Furthermore, our study did not address the robustness of our results regarding different levels of rater experience. As mentioned before, our raters had 'limited clinical experience' and were highly trained. Especially because the training is time-consuming and costly, future work could employ self-trained raters based on the extensive rater manual. Another drawback was that raters had to assess three videos from the same therapist and the same patient. Therefore, as in similar studies (Bryant, Simons, & Thase, 1999; Svartberg, 1999; von Con-sbruch et al., 2012), the data set was not independent. Thus, we could not rule out rater confounding variables. Finally, we did not test whether the raters had used the scale correctly. In order to ensure accurate scale use, Jones, Whiteside, and Neighbours (2007) suggest measuring the consistency of rater evaluations with a predetermined standard rating. In the case of our adherence and competence rating, no such gold standard was available. However, we at least ensured that the first rater performed his rating in a homogenous manner, which is illustrated by the stability over time.

Apart from these limitations, our findings suggested that the TACs-G was a reliable, appropriate, and efficient instrument for assessing competence and adherence in grief-focused CBT. The high level of reliability and the structure of the TACs-G with phase-specific items, allows for various applications in treatment process and outcome studies. For example, the scale can be employed in process research (e.g. session-by-session ratings) to identify therapy elements and associated skills which have a significant impact on symptom reduction. Moreover, the TACs-G might serve purposes that go beyond the research setting. It can identify the main therapeutic skills for a grief-focused treatment and this, in turn, can be used for the targeted teaching of these very skills. Thus, in the long run, the TACs-G might even help to improve clinical training (Perepletchikova & Kazdin, 2005). At the same time, it could

serve as a feedback basis both for trainees in supervision and certification purposes (Sholomskas et al., 2005). In summary, the TACs-G may have the potential to foster competent treatment of patients with PGD, which appears to be more important than ever considering the expected COVID-induced increase in the prevalence rate (Eisma & Boelen, 2021). Ultimately, we hope that this study will help to raise overall awareness of the importance of integrity measurements – although they are costly and resource-intensive.

Acknowledgments

The authors are very grateful to all the participants for supporting this study. RR is the author of the PG-CBT manual.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by a grant from the German Research Foundation to RR [DFG RO 2042/7-1/2]. Additionally, this work was supported by the German Research Foundation (DFG) within the funding programme Open Access Publishing.

Ethic statement

The study protocol has been approved by the IRB of the Catholic University Eichstaett- Ingolstadt (2016/21).

Data availability statement

The rating data and analysis code can be obtained J. Haneveld (Julia.haneveld@ku.de).

ORCID

Julia Haneveld  <http://orcid.org/0000-0002-3030-6145>
 Rita Rosner  <http://orcid.org/0000-0002-7960-8398>
 Anna Vogel  <http://orcid.org/0000-0001-9035-7588>
 Julia Tremel  <http://orcid.org/0000-0002-1756-5425>
 Regina Steil  <http://orcid.org/0000-0002-5367-5664>
 Winfried Rief  <http://orcid.org/0000-0002-7019-2250>
 Hannah Comtesse  <http://orcid.org/0000-0003-4150-6107>

References

- APA. (2022). *APA releases diagnostic and statistical manual of mental disorders*, Fifth Edition, Text Revision (DSM-5-TR). Retrieved 15 April 2022, from <https://www.psychiatry.org/newsroom/news-releases/apa-releases-diagnostic-and-statistical-manual-of-mental-disorders-fifth-edition-text-revision-dsm-5-tr>.
- Barber, J. P., Liese, B. S., & Abrams, M. J. (2003). Development of the cognitive therapy adherence and competence scale. *Psychotherapy Research*, 13(2), 205–221. doi:10.1093/ptr/kpg019

- Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice*, 38(5), 493–500. doi:10.1037/0735-7028.38.5.493
- Barber, J. P., Triffleman, E., & Marmar, C. (2007). Considerations in treatment integrity: Implications and recommendations for PTSD research. *Journal of Traumatic Stress*, 20(5), 793–805. doi:10.1002/jts.20295
- Bjaastad, J. F., Haugland, B. S., Fjermestad, K. W., Torsheim, T., Havik, O. E., Heiervang, E. R., & Öst, L. G. (2016). Competence and adherence scale for cognitive behavioral therapy (CAS-CBT) for anxiety disorders in youth: Psychometric properties. *Psychological Assessment*, 28(8), 908–916. doi:10.1037/pas0000230
- Boelen, P. A., de Keijser, J., van den Hout, M. A., & van den Bout, J. (2007). Treatment of complicated grief: A comparison between cognitive-behavioral therapy and supportive counseling. *Journal of Consulting and Clinical Psychology*, 75(2), 277–284. doi:10.1037/0022-006x.75.2.277
- Bryant, M. J., Simons, A. D., & Thase, M. E. (1999). Therapist skill and patient variables in homework compliance: Controlling an uncontrolled variable in cognitive therapy outcome research. *Cognitive Therapy and Research*, 23(4), 381–399. doi:10.1023/A:1018703901116
- Bryant, R. A., Kenny, L., Joscelyne, A., Rawson, N., Maccallum, F., Cahill, C., ... Nickerson, A. (2014). Treating prolonged grief disorder: A randomized clinical trial. *JAMA Psychiatry*, 71(12), 1332–1339. doi:10.1001/jamapsychiatry.2014.1600
- Bryant, R. A., Kenny, L., Joscelyne, A., Rawson, N., Maccallum, F., Cahill, C., ... Nickerson, A. (2017). Treating prolonged grief disorder: A 2-year follow-up of a randomized controlled trial. *Journal of Clinical Psychiatry*, 78(9), 1363–1368. doi:10.4088/JCP.16m10729
- Dittmann, C., Müller-Engelmann, M., Resick, P. A., Gutermann, J., Stangier, U., Priebe, K., ... Steil, R. (2017). Adherence rating scale for cognitive processing therapy - cognitive only: Analysis of psychometric properties. *Behavioural and Cognitive Psychotherapy*, 45(6), 661–670. doi:10.1017/s1352465816000679
- Djelantik, A., Smid, G. E., Kleber, R. J., & Boelen, P. A. (2017). Symptoms of prolonged grief, post-traumatic stress, and depression after loss in a Dutch community sample: A latent class analysis. *Psychiatry Research*, 247, 276–281. doi:10.1016/j.psychres.2016.11.023
- Dobson, K. S., & Singer, A. R. (2005). Definitional and practical issues in the assessment of treatment integrity. *Clinical Psychology: Science and Practice*, 12(4), 384–387. doi:10.1093/clipsy.bpi046
- Eisma, M. C., & Boelen, P. A. (2021). Commentary on: A call to action: Facing the shadow pandemic of complicated forms of grief. *OMEGA - Journal of Death and Dying*, doi:10.1177/003022282111016227
- Folke, S., Daniel, S. I. F., Gondan, M., Lunn, S., Tækker, L., & Poulsen, S. (2017). Therapist adherence is associated with outcome in cognitive-behavioral therapy for bulimia nervosa. *Psychotherapy*, 54(2), 195–200. doi:10.1037/pst0000107
- Ginzburg, D. M., Bohn, C., Höfling, V., Weck, F., Clark, D. M., & Stangier, U. (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy*, 50(12), 747–752. doi:10.1016/j.brat.2012.09.001
- Gutermann, J., Schreiber, F., Matulis, S., Stangier, U., Rosner, R., & Steil, R. (2015). Therapeutic adherence and competence scales for developmentally adapted cognitive processing therapy for adolescents with PTSD. *European journal of psychotraumatology*, 6, 26632. <https://doi.org/10.3402/ejpt.v6.26632>
- Haneveld, J., & Comtesse, H. (2020). *Rater manual for the therapeutic adherence and competence scale for grief (TACs-G)*. Katholieke Universiteit Ingolstadt- Eichstätt. Available on request.
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238–247. doi:10.1037/1040-3590.7.3.238
- Johannsen, M., Damholdt, M. F., Zachariae, R., Lundorff, M., Farver-Vestergaard, I., & O'Connor, M. (2019). Psychological interventions for grief in adults: A systematic review and meta-analysis of randomized controlled trials. *Journal of Affective Disorders*, 253, 69–86. doi:10.1016/j.jad.2019.04.065
- Jones, S., Whiteside, U., & Neighbors, C. (2007). Differences in inter-rater reliability and accuracy for a treatment adherence scale. *Cognitive Behaviour Therapy*, 36, 230–239. doi:10.1080/16506070701584367
- Lambert, M. J. (2013). *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed.). Hoboken, NJ: John Wiley.
- Loeb, K. L., Wilson, G. T., Labouvie, E., Pratt, E. M., Hayaki, J., Walsh, B. T., ... Fairburn, C. G. (2005). Therapeutic alliance and treatment adherence in two interventions for bulimia nervosa: a study of process and outcome. *Journal of Consulting and Clinical Psychology*, 73(6), 1097–1107. doi:10.1037/0022-006x.73.6.1097
- Meier, A., McGovern, M. P., Lambert-Harris, C., McLeman, B., Franklin, A., Saunders, E. C., & Xie, H. (2015). Adherence and competence in two manual-guided therapies for co-occurring substance use and posttraumatic stress disorders: clinician factors and patient outcomes. *The American Journal of Drug and Alcohol Abuse*, 41(6), 527–534. doi:10.3109/00952990.2015.1062894
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11(3), 247–266. doi:10.1016/0272-7358(91)90103-2
- Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice*, 12(4), 365–383. doi:10.1093/clipsy.bpi045
- Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, 75(6), 829–841. doi:10.1037/0022-006x.75.6.829
- Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice* (3rd ed.). Somerset, NJ: Pearson.
- Prigerson, H. G., Horowitz, M. J., Jacobs, S. C., Parkes, C. M., Aslan, M., Goodkin, K., ... Maciejewski, P. K. (2009). Prolonged grief disorder: Psychometric validation of criteria proposed for DSM-V and ICD-11. *PLOS Medicine*, 6(8), e1000121. doi:10.1371/journal.pmed.1000121
- Rosner, R., Bartl, H., Pfoh, G., Kotoučová, M., & Hagl, M. (2015). Efficacy of an integrative CBT for prolonged grief disorder: A long-term follow-up. *Journal of Affective Disorders*, 183, 106–112. doi:10.1016/j.jad.2015.04.051
- Rosner, R., Lumbeck, G., & Geissner, E. (2011). Effectiveness of an inpatient group therapy for comorbid complicated grief disorder. *Psychotherapy Research*, 21(2), 210–218. doi:10.1080/10503307.2010.545839

- Rosner, R., Pfoh, G., Kotoučová, M., & Comtesse, H. (2015). Integrative kognitive Verhaltenstherapie für die anhaltende Trauerstörung: Vorstellung eines Behandlungsmanuals. *Verhaltenstherapie*, 29(1), 40–48. doi:10.1159/000489509
- Rosner, R., Pfoh, G., Kotoučová, M., & Hagl, M. (2014). Efficacy of an outpatient treatment for prolonged grief disorder: A randomized controlled clinical trial. *Journal of Affective Disorders*, 167, 56–63. doi:10.1016/j.jad.2014.05.035
- Rosner, R., Pfoh, G., Rojas, R., Brandstätter, M., Rossi, R., Lumbeck, G., ... Geissner, E. (2015). *Anhaltende Trauerstörung*. Manual für die Einzel- und Gruppentherapie Hogrefe.
- Rosner, R., Rimane, E., Vogel, A., Rau, J., & Hagl, M. (2018). Treating prolonged grief disorder with prolonged grief-specific cognitive behavioral therapy: Study protocol for a randomized controlled trial. *Trials*, 19(1), 241. doi:10.1186/s13063-018-2618-3
- RStudio, T. (2020). *RStudio: Integrated Development for R*. Boston, MA: RStudio, PBC. <https://www.rstudio.com/>.
- Schermelleh-Engel, K., & Werner, C. S. (2012). Testtheorie und Fragebogenkonstruktion. In H. Moosbrugger, & A. Kelava (Eds.), *Methoden der Reliabilitätsbestimmung* (2nd ed., pp. 120–141). Heidelberg: Springer.
- Schlosser, R. W. (2002). On the importance of being earnest about treatment integrity. *AAC: Augmentative and Alternative Communication*, 18(1), 36–44. doi:10.1080/714043395
- Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S., ... Imber, S. D. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology*, 67(6), 837–846. doi:10.1037/0022-006x.67.6.837
- Shear, K. (2015). *Complicated Grief Treatment. Instruction Manual used in NIMH Grants*. Columbia Center for Complicated Grief.
- Shear, M. K., Reynolds, C. F., Simon, N. M., Zisook, S., Wang, Y., Mauro, C., ... Skritskaya, N. (2016). Optimizing treatment of complicated grief: A randomized clinical trial. *JAMA Psychiatry*, 73(7), 685–694. doi:10.1001/jamapsychiatry.2016.0892
- Shear, M. K., Wang, Y., Skritskaya, N., Duan, N., Mauro, C., & Ghesquiere, A. (2014). Treatment of complicated grief in elderly persons: A randomized clinical trial. *JAMA Psychiatry*, 71(11), 1287–1295. doi:10.1001/jamapsychiatry.2014.1242
- Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: A dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology*, 73(1), 106–115. doi:10.1037/0022-006X.73.1.106
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. doi:10.1037//0033-2909.86.2.420
- Svartberg, M. (1999). Therapist competence: Its temporal course, temporal stability, and determinants in short-term anxiety-provoking psychotherapy. *Journal of Clinical Psychology*, 55(10), 1313–1319. doi:10.1002/(SICI)1097-4679(199910)55:10<1313::AID-JCLP12>3.0.CO;2-F
- Verdery, A. M., Smith-Greenaway, E., Margolis, R., & Daw, J. (2020). Tracking the reach of COVID-19 kin loss with a bereavement multiplier applied to the United States. *Proceedings of the National Academy of Sciences*, 117(30), 17695–17701. doi:10.1073/pnas.2007476117
- Vogel, A., Pfoh, G., & Rosner, R. (2017). *PG13 + 9: Intervention for prolonged grief - Revised and extended translation of the PG-13*. Eichstaett: University of Eichstaett-Ingolstadt.
- von Consbruch, K., Clark, D. M., & Stangier, U. (2012). Assessing therapeutic competence in cognitive therapy for social phobia: Psychometric properties of the cognitive therapy competence scale for social phobia (CTCS-SP). *Behavioural and Cognitive Psychotherapy*, 40(2), 149–161. doi:10.1017/s1352465811000622
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, 61(4), 620–630. doi:10.1037/0022-006X.61.4.620
- Webb, C. A., Derubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 78(2), 200–211. doi:10.1037/a0018912
- Weck, F., Bohn, C., Ginzburg, D. M., & Stangier, U. (2011). Behandlungsintegrität: Implementierung, messung, evaluation und zusammenhänge zum therapieerfolg. [Treatment integrity: Implementation, assessment, evaluation, and correlations with outcome]. *Verhaltenstherapie*, 21(2), 99–107. doi:10.1159/000328840
- Weck, F., Grikscheit, F., Höfling, V., & Stangier, U. (2014). Assessing treatment integrity in cognitive-behavioral therapy: Comparing session segments with entire sessions. *Behavior Therapy*, 45(4), 541–552. doi:10.1016/j.beth.2014.03.003
- 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
- Wittchen, H. U., Zaudig, M., & Fydrich, T. (1997). SKID. Strukturiertes klinisches interview für DSM-IV. Achse I und II. Handanweisung. [SKID. Structural clinical interview of DSM-IV. Axis I and II. Manual instruction]. Hogrefe.
- World Health Organisation. (2022). *Covid-19*. Retrieved 1 March 2022 from <https://covid19.who.int>.
- Zarafonitis-Müller, S., Kuhr, K., & Bechdolf, A. (2014). Der Zusammenhang der Therapeutenkompetenz und Adhärenz zum Therapieerfolg in der Kognitiven Verhaltenstherapie—Metaanalytische Ergebnisse. [The relationship between therapist's competence and adherence to outcome in cognitive-behavioural therapy—results of a metaanalysis]. *Fortschritte der Neurologie, Psychiatrie*, 82(9), 502–510. doi:10.1055/s-0034-1366798