

The Inventory of Complicated Grief— A Systematic Psychometric Review and Conceptual Replication Study of the Structural Validity

Assessment
2023, Vol. 30(5) 1418–1434
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10731911221100980
journals.sagepub.com/home/asm



Alexander Schakowski¹, Justus Tönnies¹, Hans-Christoph Friederich¹,
Mechthild Hartmann¹, and Markus W. Haun¹ 

Abstract

The Inventory of Complicated Grief (ICG) is a commonly used self-report measure in psycho-oncology, best supportive care, and palliative medicine. However, existing validation studies yielded conflicting results regarding the structural validity. This study provides a psychometric review and conceptual replication of the ICG latent structure to test the hypothesis that existing studies overfit unreliable sources of variance, which overshadow the unidimensionality of the ICG. All proposed latent models identified in the psychometric review were tested in a series of confirmatory and exploratory structural equation models. Specifically, at least five to six latent intercorrelated factors were necessary to reach acceptable model fit. However, a general CG factor accounted for most variance and ICG sum scores showed predictable associations with anxiety and depressive symptoms, which suggests that the ICG is essentially unidimensional. There are indications that other measures of pathological grief show similar inconsistencies. Overall, potentially emerging subfacets of the ICG should not be interpreted as distinct “symptom clusters.” If time constraints are an issue as is often the case in clinical research, complicated grief may just be measured by a reduced item set without a significant loss of information or complexity.

Keywords

Inventory of Complicated Grief, COSMIN, exploratory structural equation modeling, unidimensionality

Losing someone is an exceptionally distressing event (Shear, 2015) and will cause a usually limited period of grief. However, grieving is complicated by emotional, cognitive, and behavioral maladaptations in some cases (Prigerson, Maciejewski, et al., 1995). Often resulting in severely impaired psychosocial functioning (Shear, 2015) these maladaptations serve as the diagnostic rationale for Complicated Grief (CG)¹ (Maciejewski et al., 2016; Mauro et al., 2019; Simon et al., 2020). With a prevalence of up to 11% in the general population, CG poses a substantial disease burden both on the individual and the society (Lundorff et al., 2017). One of the most commonly applied (Tremel et al., 2020) self-report measures for CG is the Inventory of Complicated Grief (ICG) (Prigerson, Maciejewski, et al., 1995). Originally, CG as measured by the ICG was conceptualized as a unidimensional construct (Jacobs et al., 2000; Prigerson, Maciejewski, et al., 1995). Although several validation studies supported the unidimensionality of the ICG (Carmassi et al., 2014; Lumbeck et al., 2012), some studies reported more complex solutions (with up to six

latent factors, Fisher et al., 2017; Simon et al., 2011) and a recent review concluded that “results varied regarding the factor structure” (Tremel et al., 2020, p. 424). Although Tremel et al. (2020) provide a general overview of measurement instruments for CG, an in-depth review of the by far most frequently used one among clinicians, the ICG, against established quality criteria (e.g., the COnsensus-based Standards for the selection of health status Measurement INstruments [COSMIN] guidelines), is lacking. Without such, the underlying causes for the inconsistent factor structure and the extent to which the interpretability of ICG scores is threatened remain unknown. Specifically, neither existing validation studies nor the recent review by Tremel

¹University Hospital Heidelberg, Germany

Corresponding Author:

Markus W. Haun, Department of General Internal Medicine and Psychosomatics, University Hospital Heidelberg, Im Neuenheimer Feld 410, Heidelberg, D-69120, Germany.
Email: markus.haun@med.uni-heidelberg.de

et al. (2020) has addressed potential causes for insufficient factorial validity (e.g., Li & Prigerson, 2016; Trembl et al., 2020) or have reasoned a posteriori that characteristics of the population under study determined the factorial structure of the ICG (e.g., Lifshitz et al., 2022). From a clinical and theoretical perspective, the assumption that different (statistical) study populations produce different latent signatures of CG seems plausible and luring. However, existing studies investigating the factor structure of the ICG did not specify hypotheses on how population effects influence factorial validity, but only reasoned such effects a posteriori which weakens the generalizability of such claims (Deffner et al., 2021). Epistemologically, the finding that the factorial structure of the ICG varies as the population under study varies is a mere correlational finding and as such only a sufficient condition for a *causal* population effect. The necessary condition for such a causal effect, that is, keeping the population under study constant should not affect the factorial structure of the ICG, has not been tested, yet. If it turns out, as we will show below, that this necessary condition is not met, a different explanation for the insufficient structural validity will be needed. Here, we argue that existing studies on structural validity overfitted unsystematic variance in latent measurement models and underestimated the extent to which a general CG factor accounts for reliable variance (Reise, Bonifay, et al., 2013; Rodriguez et al., 2016). Whereas a lack of unidimensionality would pose a conceptual threat to CG, in practice, ICG total scores might be readily interpretable and might reliably predict other constructs (Luo & Al-Harbi, 2016; Reise et al., 2007; Reise, Bonifay, et al., 2013; Rodriguez et al., 2016). The essential unidimensionality of the ICG would offer a simple explanation for the disagreement on the number and content of ICG subfacets in the existing literature without assuming that a potentially unlimited number of population-specific variables have strong influences on the factorial structure of the ICG. In addition, clinical researchers or clinicians could interpret ICG sum scores without having to assume that the expression of CG drastically varies from one patient population to another.

This systematic psychometric review and conceptual replication study therefore (a) systematically reviews and benchmarks the measurement properties of the ICG with a focus on structural validity against well-established and easily interpretable quality criteria, that is, the COSMIN (Mokkink, de Vet, et al., 2018) to gain a more detailed psychometric assessment of the ICG than the existing review by Trembl et al. (2020) offered; (b) evaluates and compares proposed measurement models for the ICG and assesses the degree of multidimensionality in ICG items; and (c) evaluates whether the ICG is essentially unidimensional. The overarching aim is to provide a psychometric synopsis of the ICG relevant both to clinical practice and theoretical research.

Method

Study Design and Recruitment

First, we conducted a systematic review of studies reporting measurement properties of the ICG following the COSMIN standards (Mokkink, Prinsen, et al., 2018). Second, to assess multidimensionality, we conducted confirmatory factor analysis (CFA)/bifactor-CFA and exploratory structural equation modeling (ESEM)/bifactor-ESEM analyses to disentangle different sources of construct variance (Morin, Arens, & Marsh, 2016; Morin, Arens, Tran, et al., 2016) in two independent samples of caregivers of bereaved German people with cancer (Sample 1 and replication Sample 2). Third, we assessed whether ICG scores are essentially unidimensional (Reise, Bonifay, et al., 2013; Rodriguez et al., 2016) and provide further evidence for some of the psychometric properties of the ICG. Both samples were approved by the Institutional Review Board of the Medical Faculty of Heidelberg University. We report all data exclusions, all manipulations, and all measures in the study.

Sampling Procedure for Sample 1. For Sample 1, we identified bereaved caregivers of cancer patients from the Cancer Registry of the National Center for Tumor Diseases Heidelberg (NCT) and administered paper pencil surveys via mail. Tumor types were restricted to colon, pancreatic, lung, breast, and prostate cancer. All adult caregivers who were able to give informed consent were eligible. Exclusion criteria comprised patient's death not longer than 6 months ago and enrolment of the bereaved caregivers in a cancer treatment trial. A more detailed description of the sampling procedure can be found in Tönnies et al. (2021).

Sampling Procedure for Sample 2. For Sample 2, we identified 21 German grief support websites and online groups in a systematic online search and invited caregivers of deceased individuals to fill in an online survey. We also invited participants to take part in a follow-up survey. A detailed description of the sampling procedure can be found in Haun et al. (2019).

Participants in Sample 1. From 1,138 patients who were detected in the Cancer Registry of the NCT, 646 bereaved caregivers were identified as potentially eligible. In total, 298 caregivers participated in the survey (response rate: 46.1%). Accounting for multivariate outliers on ICG items (Mahalanobis D^2) and missing data for the ICG total score, 259 datasets were analyzed (cf. Tönnies et al., 2021). Table 1 shows the participant characteristics. Histograms of ICG scores, along with estimated polychoric correlations between ICG items can be found in Online Appendix C.

Table 1. Sample Characteristics for Both Samples.

Characteristics	Sample 1 (N = 259)			Sample 2 (N = 359)			Comparison ^a p
	N ^b (%)	M (SD)	Median Skewness	N (%)	M (SD)	Median Skewness	
Age in years		60.89 (14.94)	63 -0.58		46.8 (10.9)	48 -0.24	<.001
Gender							<.001
Female	153 (59.5)			324 (90.3)			
Male	104 (40.5)			35 (9.7)			
Relationship status							<.001
Married/de facto/living together	75 (29.8)			214 (59.3)			
Single	177 (70.2)			139 (38.5)			
Employment status							<.001
Retired	135 (52.7)			28 (7.8)			
Full time/self-employed/student/domestic work	101 (39.5)			296 (82.5)			
Unemployed	7 (2.7)			14 (3.9)			
Other	13 (5.1)			21 (5.8)			
I am the . . . of the deceased							<.001
Spouse/partner	177 (69.4)			133 (37.0)			
Child/stepchild	64 (25.1)			131 (36.5)			
Parent	4 (1.6)			28 (7.8)			
Sibling	3 (1.2)			24 (6.7)			
Other/not specified	4 (1.6)			43 (12.0)			
Age of deceased in years		66.7 (10.61)	68 -0.40		59.7 (14.20)	60 -0.20	<.001
Time between initial tumor diagnosis and death of deceased in months (= 30 d)		29.44 (37.07)	16.95 3.26		33.5 (42.7)	18.0 2.8	.220
Time between death of deceased and survey participation in months		27.09 (9.71)	27.07 2.16		43.6 (48.4)	27.0 2.8	<.001
ICG total score		27.19 (14.24)	25 0.43		27.84 (13.76)	26 0.39	.573
ICG total score > 30	104 (40.15)				149 (41.27)		.843
PHQ-9 total score		9.51 (6.18)	8 0.46		8.48 (6.71)	7 0.70	.050
GAD-7 total score		7.41 (5.23)	6 0.68		7.39 (5.39)	6 0.67	.961

Note. M = mean; SD = standard deviation; ICG = Inventory Of Complicated Grief; PHQ-9 = Patient Health Questionnaire; GAD-7 = Generalized Anxiety Disorder Scale.

^aFor metric outcomes, independent t tests for unequal variances were conducted; for categorical outcomes, χ^2 tests were conducted.

^bVariables containing missing data do not add to the total sample size.

Participants in Sample 2. In total, 559 individuals were eligible of whom 365 participated in the survey (response rate: 65.3 %). Accounting for multivariate outliers on ICG items (Mahalanobis D^2) and after excluding two cases with missing data on all ICG variables, 359 datasets were analyzed in this replication study (cf. Haun et al., 2019). Table 1 shows the participant characteristics. Sixty participants filled in the retest survey. The median time interval between test and retest was 12 weeks. Histograms of ICG scores, along with estimated polychoric correlations between ICG items can be found in Online Appendix C.

Measures for Sample 1 and Sample 2

Inventory of Complicated Grief. The ICG consists of 19 items (Lumbeck et al., 2012). Individuals describe the currently experienced emotional, cognitive, and behavioral states on a 5-point unipolar frequency scale from “0” (*never*) to “4” (*always*).

Patient Health Questionnaire–9. The PHQ-9 is a widely used brief depression severity measure with high validity and reliability which scores each of the nine *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; American Psychiatric Association, 1994) criteria as “0” (*not at all*) to “3” (*nearly every day*) (Kroenke et al., 2001).

Generalized Anxiety Disorder–7. The seven-item Generalized Anxiety Disorder–7 (GAD-7) is a valid self-report anxiety questionnaire with unidimensional structure and good internal consistency (Hinz et al., 2017). On a 4-point unipolar frequency scale, individuals indicate how often they have experienced symptoms of generalized anxiety during the last 2 weeks.

Part 1: Psychometric Review

We conducted a review of studies reporting any measurement property of the ICG comprised in the COSMIN standards except content validity (Mokkink, de Vet, et al., 2018). Four electronic databases (MEDLINE, PsycInfo, Web of Science, CINAHL) were searched until February 11, 2022 using instrument name and measurement properties as search terms (Terwee et al., 2009) (Online Appendix A). All studies reporting any COSMIN measurement property in any study population were included. We excluded studies that merely reported correlations of ICG scores with other constructs or studies that reported measurement properties of instruments adapted from the original ICG (e.g., the ICG-revised, O’Connor et al., 2010). Two reviewers (AS and MWH) independently screened titles and abstracts of retrieved articles for relevance. Evidence for measurement properties was retrieved independently by two reviewers (AS and JT) and assessed using the COSMIN criteria. For individual studies,

evidence on each measurement property is rated from “inadequate,” “doubtful,” “adequate” to “very good.” Evidence on measurement properties is then synthesized across studies as sufficient (+), insufficient (–), inconsistent (+/–), or indeterminate (?). The quality of evidence of each measurement property was rated “high,” “moderate,” “low,” or “very low” according to the Grading of Recommendations Assessment, Development, and Evaluation approach (Mokkink, Prinsen, et al., 2018; Terwee et al., 2012). For a more detailed description, see Online Appendix A.

Part 2: Multidimensionality

To assess the sources of multidimensionality of the ICG, we conducted a series of (Bifactor-) CFA and (Bifactor-) ESEM analyses (Morin, Arens, & Marsh, 2016; Morin, Arens, Tran, et al., 2016) for all latent models of the ICG identified in the psychometric review in both samples, that is Sample 1 and Sample 2. Analyses were carried out using Mplus v8.4 (Muthén & Muthén, 2019). First, we conducted CFA and ESEM analyses with oblique target rotation. Second, we estimated orthogonal Bifactor-CFA and orthogonal Bifactor-ESEM models with target rotation in both samples to disentangle the amount of variance that is explained by one general CG factor and orthogonal group factors. Group factors correspond to the proposed latent factors of the ICG. For target rotation, we considered the target to be reached if items loaded most highly on the specified factor (structural replicability, Osborne & Fitzpatrick, 2012, see also Litalien et al., 2017). To account for the Likert-type scale of the ICG, all parameters were estimated using Weighted-Least-Squares estimation with mean and variance adjustment (WLSMV). For model evaluation, we inspected the statistical significance of model parameters along with goodness-of-fit indices (χ^2 , comparative fit index, comparative fit index [CFI], standardized root-mean-square residual [SRMR], and root-mean-square error of approximation [RMSEA]). As commonly applied criteria to evaluate model fit (e.g., Hu & Bentler, 1999) based on maximum-likelihood (ML) estimated models are biased (i.e., increased probability of accepting misspecified models) when applied to WLSMV estimated models (Nye & Drasgow, 2011; Shi & Maydeu-Olivares, 2020; Shi et al., 2020; Xia & Yang, 2019), we obtained approximate sampling distributions of model fit indices if models are correctly specified and misspecified using Monte-Carlo-Simulations (see Online Appendix D for details) in R 4.0.2 (R Core Team, 2020). In particular, we repeatedly simulated data from a “true” model and fitted either correctly specified or incorrectly specified SEM models to this data using WLSMV and recorded CFI, RMSEA, and SRMR. The resulting distributions were then used to obtain cut-off values that minimize Type-II-error (β) to detect moderate model misspecification (three to five

moderately intercorrelated residuals that were not specified in the fitted model) (Niemand & Mai, 2018). We assumed good model fit for CFI > .977, RMSEA < .038 and SRMR < .061 ($1 - \beta = .90$) and adequate model fit for CFI > .972, RMSEA < .043 and SRMR < .063 ($1 - \beta = .80$). For these values less than 5% of correctly specified models were rejected ($\alpha < .05$). In addition, we calculated information criteria (Bayesian information criterion [BIC] and Akaike information criterion [AIC]) for each model based on the minimum of the WLSMV discrepancy function (Huang, 2017).

Part 3: Unidimensionality and Psychometric Properties

Unidimensionality. Omega coefficients derived from bifactor models allow us to assess essential unidimensionality (Reise, Scheines, et al., 2013; Rodriguez et al., 2016). For reliability, we report coefficient Ω that can be interpreted equivalently to Cronbach's α , coefficient Ω_S for all subscales (reliability of a composite measure of the general factor and the group factor S), and coefficient Ω_H (relative amount of variance explained by the general factor H). In addition, we provide coefficients Ω_{H-S} (reliability of subscales S with the general factor H partialled out). To further investigate essential unidimensionality, we report explained common variance (ECV) coefficients, which quantify the variance explained by the general factor relative to the variance explained by the general factor and all group factors. To estimate the extent to which latent ICG factors are reproducible and consistent, we compute factor replicability coefficients, H, which are a function of the ratio of latent variance explained within each latent factor relative to the variance not explained within each latent factor. All coefficients are reported for the best fitting Bifactor CFA models in both samples. Following Rodriguez et al. (2016), we conclude that the ICG is essentially unidimensional if Ω , Ω_S , $\Omega_H > .80$ and $ECV > .70$, while Ω_{H-S} is low for all subscales S. We conclude that latent factors do not constitute reproducible latent subfacets of the ICG if factor replicability $H < .80$ (Rodriguez et al., 2016).

Psychometric Properties. Measurement properties of the ICG are analyzed in accordance with the COSMIN guidelines and the COSMIN study design checklist (Mokkink et al., 2019).

Construct Validity. To illustrate our results on unidimensionality, we compared correlations between comparator constructs (depression and anxiety as measured by the PHQ-9 and GAD-7) and ICG total scores extracted from a well-fitting multidimensional solution with correlations between comparator constructs and ICG total scores

extracted from the unidimensional solution in both samples. We hypothesized ICG scores to be moderately correlated with PHQ-9 scores and GAD-7 scores (e.g., Lumbeck et al., 2012; Prigerson, Frank, et al., 1995), irrespective of the factor extraction method. To estimate these correlations, we extracted factor scores from unidimensional CFA models for ICG, PHQ-9 and GAD-7 and from the best fitting Bifactor ESEM model of the ICG (Litalien et al., 2017; Morin, Meyer, et al., 2016; Skrandal & Laake, 2001) in both samples.

Reliability and Measurement Error. To assess reliability and measurement error, we analyzed the subset of the web survey sample that completed the ICG twice ($n = 60$). We assumed that ICG scores were stable in the interim period. We computed the standard error of measurement (SEM) along with the smallest detectable change (SDC). For test-retest reliability we calculated the intraclass correlation coefficient (ICC; A,1; two-way mixed-effects model with absolute agreement).

Responsiveness. Following an exploratory approach, we analyzed associations between GAD-7, PHQ-9, and ICG change scores between T1 and T2 in the test-retest subset of Sample 2 ($n = 60$). Specifically, we tested whether depressive or anxiety symptoms at T1 were associated with ICG change scores between T1 and T2.

Results

Part 1: Psychometric Review

Figure 1 shows a flowchart of the review process. We identified full-text articles for 13 studies published between 1995 and 2020 that report measurement properties of the ICG. Of those, 11 studies were included in the review. Table 2 summarizes the existing evidence on measurement properties. For a detailed description of reported results see Table A3 in Online Appendix A. We found that, except for internal consistency (nine studies of very good quality), criterion validity (one study of very good quality),² and construct validity (seven studies of very good quality),³ there is only weak evidence to support each of the remaining measurement properties of the ICG. There was no study reporting on responsiveness or measurement error. Only two studies provided evidence of moderate quality (downgraded for inconsistency) for measurement invariance with respect to sample type (clinical vs. non-clinical, Fisher et al., 2017) and for differential item functioning between individuals scoring above 25 on the ICG and those scoring below (Masferrer et al., 2017). Most strikingly, we found that evidence on structural validity of the ICG is highly inconsistent, stemming from studies that applied non-optimal methods (only one study reached the level of very good

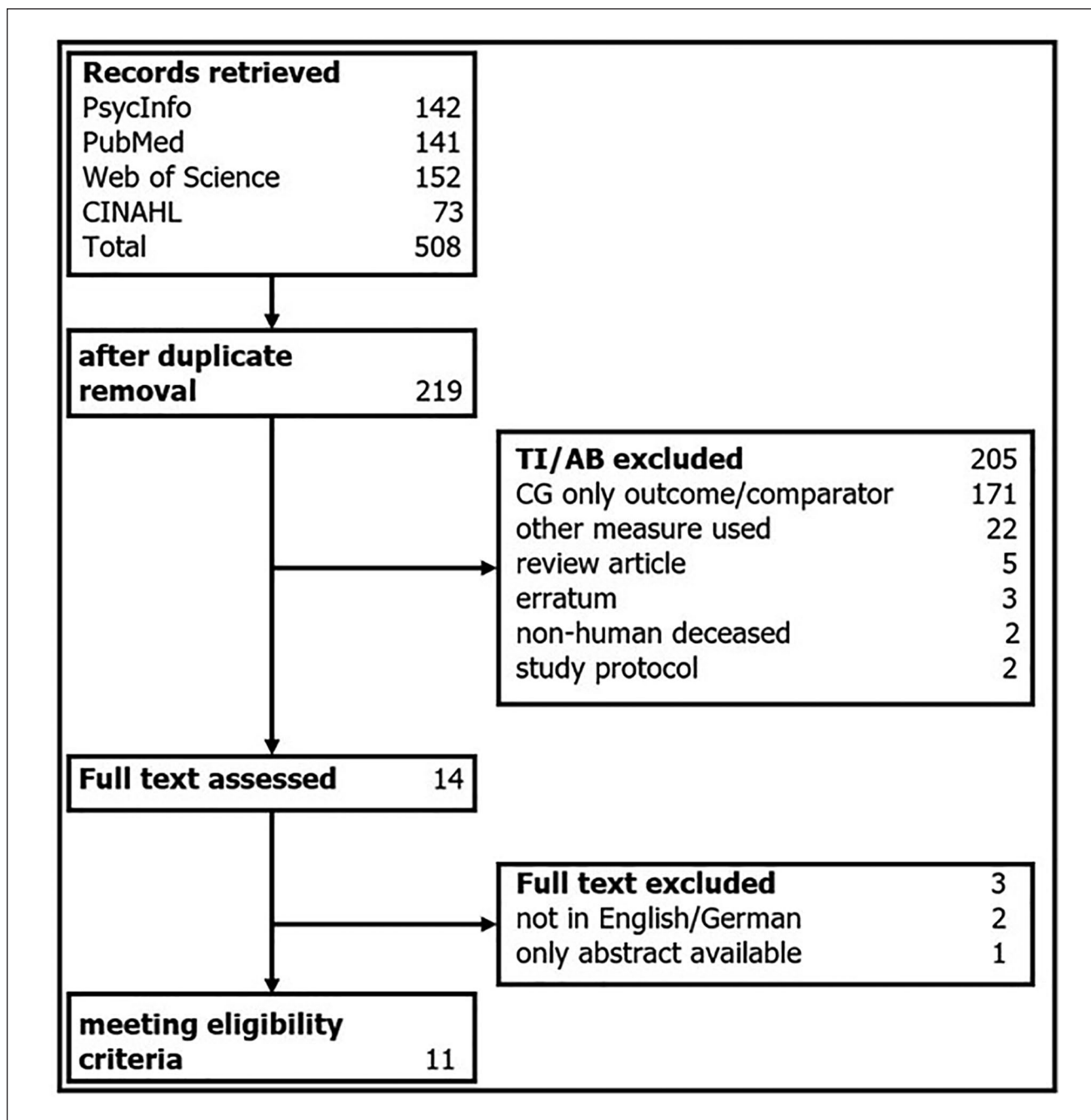


Figure 1. Flowchart Displaying the Process of Study Selection.

quality, but overall quality of evidence regarding structural validity had to be downgraded to low due to inconsistency). We identified eight different factor models proposed to capture the latent structure of the ICG, ranging from unidimensional models (e.g., Prigerson, Maciejewski, et al., 1995, 5/11 studies), three-factor models (3/11 studies) with factor intercorrelations (Thimm et al., 2019) and without factor intercorrelations (Li & Prigerson, 2016; Lifshitz et al.,

2022), four-factor models (Masferrer et al., 2017), to six-factor models with factor intercorrelations (Fisher et al., 2017) and without factor intercorrelations (Simon et al., 2011). Most studies applied exploratory factor analysis (EFA)/principal components analysis (PCA) (e.g., Lifshitz et al., 2022; Lumbeck et al., 2012, 7/11 studies). Most EFA/PCAs were based on Pearson correlations (4/7 studies). Three studies (Carmassi et al., 2014; Masferrer et al., 2017;

Table 2. Evidence Synthesis.

Reference	Prigerson, Maciejewski, et al. (1995)	Lumbeck et al. (2012)	Ludwikowska-Świeboda & Lachowska (2019)	Carmassi et al. (2014)	Simon et al. (2011)	Li & Prigerson (2016)	Lifshitz et al. (2022)	Masferrer et al. (2017)	Fisher et al. (2017)	Han et al. (2016)	Thimm et al. (2019)	Overall rating	
	Rating (+/-/?)	Rating (+/-/?)	Rating (+/-/?)	Rating (+/-/?)	Rating (+/-/?)	Rating (+/-/?)	Rating (+/-/?)	Rating (+/-/?)	Rating (+/-/?)	Rating (+/-/?)	Rating (+/-/?)	Overall rating (+/-/?)	Quality of evidence (high, moderate, low, very low)
Structural validity	?	?	+	+	-	+	?	?	+	?	+	+/-	Low
Internal consistency	+	+	+	+	NA	+	+	+	NA	+	+	+	High
Cross-cultural validity	NA	NA	NA	NA	NA	NA	NA	-	+	NA	NA	+/-	Moderate
Reliability	?	?	NA	NA	NA	NA	NA	NA	NA	?	NA	?	Low
Measurement error	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Criterion validity	NA	NA	NA	+	+	+	NA	NA	NA	NA	NA	+	High
Construct validity	+	+	+	+	NA	+	+	+	NA	+	+	+	High
Responsiveness	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Note. NA = not applicable.

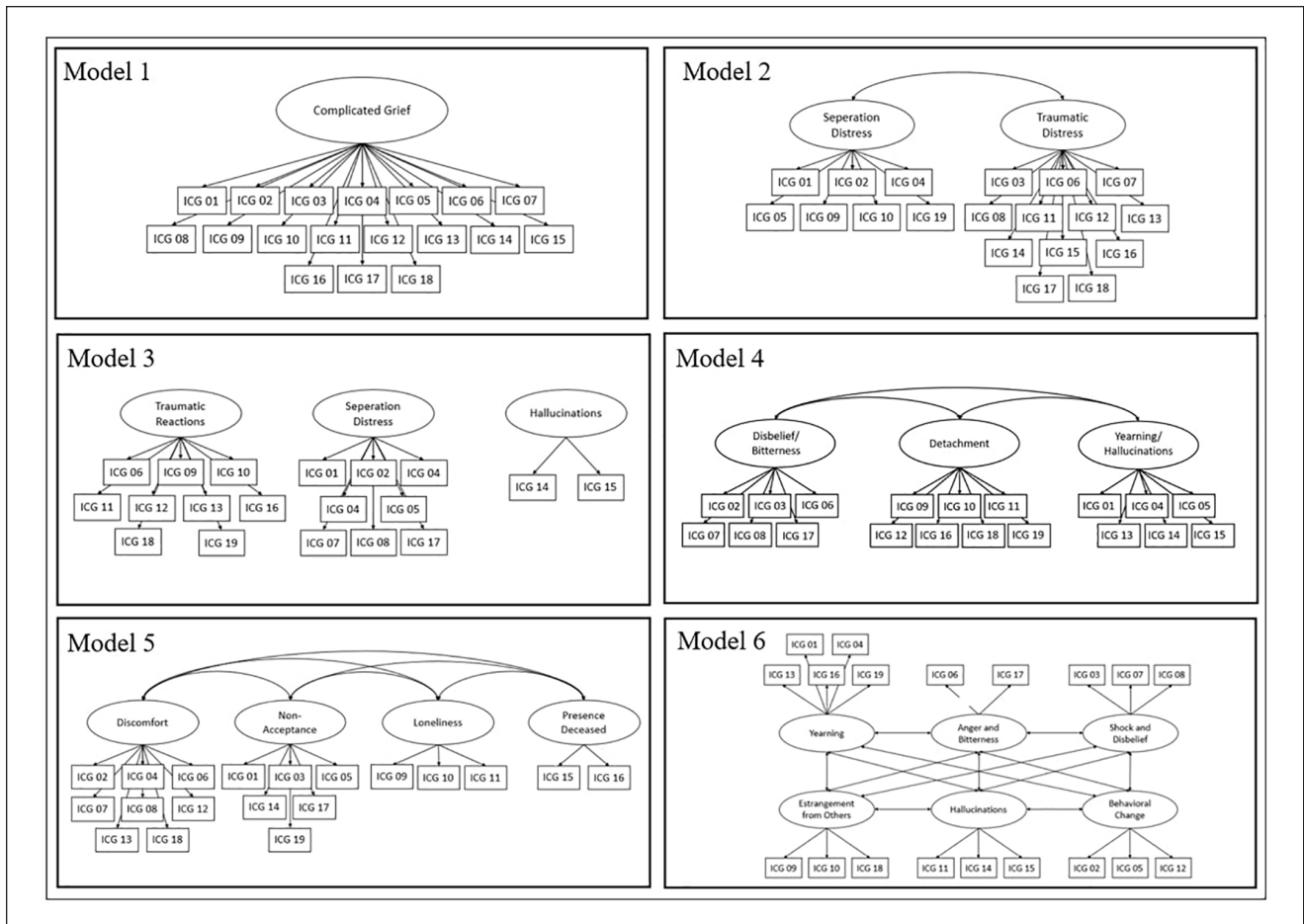


Figure 2. Latent Factor Models Identified in the Psychometric Review and Eligible for Replication.

Note. Model 1 = unidimensional model (Prigerson, Maciejewski, et al., 1995); Model 2 = two-factor model with (theoretical) dimensions Traumatic Distress and Separation Anxiety (Jacobs et al., 2000; O'Connor et al., 2010), Model 3 = three-factor model (Lifshitz et al., 2022); Model 4 = three-factor model (Thimm et al., 2019); Model 5 = four-factor model (Masferrer et al., 2017); Model 6 = six-factor model (Simon et al., 2011). Item indicators refer to (Lumbeck et al., 2012).

Simon et al., 2011) applied EFA on tetrachoric correlation matrices based on dichotomized items. None of the studies accounted for ordinal data. Five studies applied confirmatory analyses (either CFA or EFA/CFA combined).

Part 2: Multidimensionality

Of the eight identified models, one model was based on only 17 items (Lifshitz et al., 2022), and one model (Fisher et al., 2017) was based on 16 items, leaving five models eligible for comparison within both samples. In addition, we tested a two-factor solution based on the conceptual distinction between traumatic grief and separation distress discussed in the literature (e.g., Jacobs et al., 2000; O'Connor et al., 2010). Because the two-factor solution had only been tested with the ICG-r (O'Connor et al., 2010), we assigned items to each factor based on eyeball validity. For testing all models, we assumed factors to be

correlated. Final models are displayed in Figure 2. All standardized parameter estimates for all models in both samples reported below are displayed in Online Appendix B. In a first step, we tested for convergence and whether implicated covariance matrices were positive definite. In Sample 1 all CFA models converged, but Models 5 and 6 did not produce positive definite covariance matrices. Inspection yielded that the two factors discomfort and non-acceptance of model 5 as well as the two factors Yearning and Behavioral Change of Model 6 were linearly dependent ($r_s > .95$). To subject these models to CFA, we collapsed the respective factors⁴ (resulting in Model 5b and Model 6b). Table 3 shows CFA results in Sample 1. All models failed to reach adequate model fit. Although in Sample 2, Model 5b did not produce a positive definite model implied covariance matrix, we could replicate the results in general (Table B1 in Online Appendix B). That is, none of the models showed adequate model fit. Turning

Table 3. Results of CFA/Bifactor ESEM/Bifactor-ESEM of Latent ICG Models (Sample 1, N = 259).

CFA	Model	$\chi^2(df)$	Number of free parameters	CFI	RMSEA [90% CI]	SRMR	BIC	AIC
	1	504.693* (152)	95	.947	.095 [.086, .104]	.066	2.754	1.450
	2	487.963* (151)	96	.949	.093 [.084, .102]	.066	2.747	1.429
	3	450.577* (149)	98	.955	.088 [.079, .098]	.063	2.722	1.376
	4	406.352* (149)	98	.961	.082 [.072, .091]	.061	2.652	1.307
	5	—	—	—	—	—	—	—
	5b	420.000* (151)	98	.959	.084 [.074, .093]	.057	2.674	1.328
	6	—	—	—	—	—	—	—
	6b	321.719* (142)	105	.973	.070 [.060, .080]	.055	2.646	1.204
Bifactor-CFA	Model	$\chi^2(df)$	Number of free parameters	CFI	RMSEA [90% CI]	SRMR	BIC	AIC
	2	—	—	—	—	—	—	—
	3	361.796* (134)	113	.966	.081 [.071, .091]	.055	2.867	1.315
	4	289.784* (133)	114	.976	.067 [.057, .078]	.051	2.787	1.222
	5	—	—	—	—	—	—	—
	5b	—	—	—	—	—	—	—
	6	—	—	—	—	—	—	—
	6b	—	—	—	—	—	—	—
ESEM	Model	$\chi^2(df)$	Number of free parameters	CFI	RMSEA [90% CI]	SRMR	BIC	AIC
	2	382.537* (134)	113	.963	.085 [.075, .095]	.050	2.856	1.304
	3	236.275* (117)	130	.982	.063 [.051, .074]	.037	2.989	1.203
	4	236.275* (117)	130	.982	.063 [.051, .074]	.037	2.989	1.203
	5	176.374* (101)	146	.989	.054 [.040, .067]	.029	3.260	1.254
	6	104.475* (72)	175	.995	.042 [.022, .058]	.022	3.814	1.411
Bifactor-ESEM	Model	$\chi^2(df)$	Number of free parameters	CFI	RMSEA [90% CI]	SRMR	BIC	AIC
	2	236.275* (117)	130	.982	.063 [.051, .074]	.037	2.989	1.204
	3	176.374* (101)	146	.989	.054 [.040, .067]	.029	3.260	1.255
	4	176.374* (101)	146	.989	.054 [.040, .067]	.029	3.260	1.255
	5	135.907* (86)	161	.992	.047 [.031, .062]	.026	3.542	1.331
	6	—	—	—	—	—	—	—
	6b	104.475* (72)	175	.995	.042 [.022, .058]	.022	3.814	1.411

Note. CFA = confirmatory factor analysis; ESEM = exploratory structural equation modeling; ICG = Inventory of Complicated Grief; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; BIC = Bayesian information criterion; AIC = Akaike information criterion. * $p < .001$.

to Bifactor-CFA, in Sample 1, only Bifactor-Models 3 and 4 converged and implied positive definite covariance matrices but showed inadequate model fit (Table 3). This was replicated in Sample 2, where only Model 4 converged with inadequate model fit (Table B1). In sum, both CFA and Bifactor-CFA models did not capture the latent structure of the ICG to a satisfactory extent. Next, we conducted ESEM analyses for all models, to assess whether unspecified cross-loadings between factors explain the inadequate fit (Asparouhov & Muthén, 2009). For Sample 1, all models showed better fit (Table 3). However, this came at the cost of increased model complexity as indicated by higher values for BIC and AIC. Of the ESEM models, Model 6 showed close to adequate model fit. This result was replicated in Sample 2, where Model 6 even reached close to good model fit (Table B2). Note further, that factor intercorrelations were reduced dramatically for all models (e.g., from $r = .943$ to $r = .411$ in the two factor model; see Table B13 in Online Appendix B), which indicated that cross-loadings were present (Morin, Arens, & Marsh, 2016). However, we find that neither in Sample 1 nor in Sample 2, the target model was reached with three items in each study loading higher on a non-target factor (Tables B9 and B10 in Online Appendix B). Moreover, in each study, different items failed to meet the target (Items 1, 5, and 12 in Sample 1 and Items 2, 11, and 17 in Sample 2). Next, we subjected all models to Bifactor-ESEM with orthogonal target rotation. We found that in Sample 1, while the covariance matrix implied by Model 6 was not positive definite, Model 6b showed the best model fit⁵ (Table 3). In Sample 2, this result was replicated (Table B2 in Online Appendix B). Again, the target was not met with five items diverging from the target in Sample 1 and seven items in Sample 2 (Table B9 and B10 in Online Appendix B). Again, different items did not meet the target (Items 1, 2, 7, 8, 12 in Sample 1 and Items 1, 2, 12, 13, 16, 18, and 19 in Sample 2). In summary, none of the proposed latent models could be replicated in either study. To better interpret the factor loadings of the well-fitting Bifactor-ESEM Model 6b (five group factors plus one general factor), we fitted a five-factor ESEM model and a 5+1-factor Bifactor-ESEM model in Sample 1 using orthogonal geomin rotation. We specified significant loadings (see Table B11) as target loadings to be reached by the same models fitted with target rotation in Sample 2. For ESEM, we found that 11 items and for Bifactor-ESEM three items diverged from this target (Table B12). In sum, the 5- and 5+1-ESEM and Bifactor-ESEM models provided acceptable/good fit to the data but did not produce loading patterns that were replicable between studies. Nevertheless, the acceptable model fit made it possible to use these models to assess essential unidimensionality.

Part 3: Unidimensionality and Psychometric Properties

Unidimensionality. All Ω coefficients derived from the 5+1 Bifactor ESEM model are displayed in Table 4. As can be seen, Ω and Ω_S for all factors were high (all $> .96$ in both samples), indicating that composite scores of general and specific factors were reliable. More importantly, Ω_H was close to or above $.80$ in both samples indicating that ICG scores were essentially unidimensional. However, Ω_{H-S} was very low for all subscales indicating that subscales did not account for a significant amount of variance relative to the general factor. In line with this observation, ECV was above or close to $.70$ in both samples, indicating that most variance was explained by the general factor in relation to the latent group factors. Factor replicability was close to 1 for the general factor in both samples and way below $.80$ for the group factors, except Factors 2 and 3. Here, factor replicability was above and close to $.80$, respectively. Although there were only two significant non-target cross-loadings for Factor 2 in Sample 2 (Table B11 in Online Appendix B), there were six significant non-target cross-loadings for Factor 3 in Sample 2 (Table B12 in Online Appendix B) indicating that of all group factors, only Factor 2 was replicable between studies to some extent.

Construct Validity. Although in Sample 1 there was no missing data on the PHQ-9 or GAD-7, we excluded 12 participants with missing values on PHQ-9 and/or GAD-7 listwise, resulting in a total sample of $n = 347$ for Sample 2. We fitted unidimensional CFA models to measure PHQ-9, Sample 1: $\chi^2(df) = 87.339(27)$, RMSEA = $.093[.072, .115]$, CFI = $.978$, SRMR = $.049$; Sample 2: $\chi^2(df) = 89.419(27)$, RMSEA = $.082[.063, .101]$, CFI = $.992$, SRMR = $.031$, and GAD-7, Sample 1: $\chi^2(df) = 43.487(14)$, RMSEA = $.090[.061, .121]$, CFI = $.991$, SRMR = $.030$; Sample 2: $\chi^2(df) = 67.270(14)$, RMSEA = $.105[.080, .130]$, CFI = $.988$, SRMR = $.029$, and extracted latent factor scores. For ICG, we extracted latent factor scores from the unidimensional CFA and from the 5+1 Bifactor-ESEM model. In both samples, Bifactor-ESEM ICG scores were moderately correlated with CFA PHQ-9 scores, Sample 1: $r(257) = .574$, $t = 11.237$, $p < .001$, Sample 2: $r(345) = .682$, $t = 17.309$, $p < .001$, and CFA GAD-7 scores, Sample 1: $r(257) = .454$, $t = 8.172$, $p < .001$, Sample 2: $r(345) = .589$, $t = 13.536$, $p < .001$. These results did not change when using CFA ICG scores both for CFA PHQ-9 scores, Sample 1: $r(257) = .584$, $t = 11.518$, $p < .001$, Sample 2: $r(345) = .688$, $t = 17.613$, $p < .001$, and CFA GAD-7 scores, Sample 1: $r(257) = .464$, $t = 8.388$, $p < .001$, Sample 2: $r(345) = .602$, $t = 13.998$, $p < .001$. Most importantly, CFA ICG scores were almost perfectly correlated with Bifactor ESEM ICG scores

Table 4. Bifactor-Model Derived Omega and Hierarchical Omega Coefficients in Both Samples.

Coefficient		Sample 1	Sample 2
Omega coefficient	Ω	.968	.965
	Ω_{F1}	.963	.959
	Ω_{F2}	.964	.959
	Ω_{F3}	.964	.960
	Ω_{F4}	.964	.960
	Ω_{F5}	.964	.960
Hierarchical Omega coefficient	Ω_H	.761	.802
	Ω_{H-F1}	.027	.032
	Ω_{H-F2}	.056	.050
	Ω_{H-F3}	.036	.049
	Ω_{H-F4}	.034	.046
	Ω_{H-F5}	.039	.032
Explained common variance	ECV	.709	.684
Factor replicability coefficients	H_G	.979	.976
	H_{F1}	.673	.681
	H_{F2}	.894	.799
	H_{F3}	.790	.774
	H_{F4}	.636	.694
	H_{F5}	.643	.644

Note. ECV = explained common variance.

in both samples, Sample 1: $r(257) = .987, t = 97.922, p < .001$, Sample 2: $r(345) = .990, t = 131.63, p < .001$.

Responsiveness. Higher levels of depressive and anxiety symptoms at T1 were not associated with ICG change scores between T1 and T2, $r = .075[-.182, .323], t(58) = .575, p = .567$, and $r = .123 [-.132, .368], t(58) = .965, p = .339$, respectively.

Reliability and Measurement Error. The median time interval between test and retest amounted to 12 weeks (interquartile range: 6 weeks), which we considered long enough to prevent recall bias. The median proportion of weeks between test and retest of weeks since bereavement was 60% (interquartile range: 73%). The ICC amounts to $ICC(A,1) = .908$, 95% confidence interval (CI) = [.85, .94]. The SDC in ICG scores derived from the SEM of 1.853 ($n = 60$) amounts to $SDC_{ind} = 5.135$ for individuals, and $SDC_{group} = .663$ for comparisons between groups. In our sample, 22 individuals (36.7%) showed change scores between T1 and T2 that exceeded SDC_{ind} . Bland-Altman analysis indicated no systematic bias between both administrations, bias = -1.15 , 95% CI = $[-2.72, 0.42]$.

Discussion

The present study is the first to (a) provide an in-depth, and easily interpretable psychometric review of the ICG by synthesizing evidence on measurement properties using the

COSMIN standards, (b) to systematically test and compare proposed factor models for the ICG based on the psychometric review to assess the degree of multidimensionality in ICG items, and (c) to estimate the degree to which multidimensionality affects the interpretation of ICG as an essentially unidimensional construct. Regarding the psychometric soundness of the ICG, we found that despite the large use of the ICG in longitudinal studies (Milic et al., 2019; Wågø et al., 2017), there is no evidence on the responsiveness and the measurement error of the ICG, and only weak evidence on its reliability. Although existing validation studies as well as interventional studies using the ICG rely on highly heterogeneous samples (cf. Trembl et al., 2020), there is only inconsistent evidence on its measurement invariance. Notably, we partially closed this gap by providing estimates of measurement error and ICC. However, the present study focused on the structural validity of the ICG. In this regard, in line with Trembl et al. (2020), we find that results are highly inconsistent with proposed latent ICG models ranging from unidimensional to six-dimensional solutions. In total, we identified six distinct and testable latent models reported in the literature. Based on both confirmatory and exploratory latent modeling approaches, we find that assumptions of the independent clusters model CFA are too strict to be applied to factor models proposed for the ICG. This is likely due to the fact that the plethora of studies included in this review rely on EFA approaches (Marsh et al., 2010). One may argue that our findings contradict other results from earlier CFA studies on the ICG. For example, Fisher et al. (2017) reported good model fit

for the six-factor model in a CFA analysis as did Ludwikowska-Świeboda and Lachowska (2019) for a unidimensional CFA solution. However, Fisher et al. (2017) applied commonly used ML cutoffs to WLSMV estimated models potentially leading to biased conclusions (Nye & Drasgow, 2011; Shi & Maydeu-Olivares, 2020). Furthermore, the fact that several residual correlations had to be freed indicates misspecification. Similarly, Ludwikowska-Świeboda and Lachowska (2019) specified residual correlations corresponding to the six factors reported in Simon et al. (2011) as well as additional residual correlations. In sum, these models might not at all reflect unidimensionality but rather several minor factors (cf. Brown, 2015) necessary to achieve acceptable model fit. In our data, systematic improvement in model fit was instead achieved by allowing for cross-loadings in the ESEM framework. But even when allowing for cross-loadings, only a five-factor solution (Sample 1) and a six-factor solution (Sample 2) reached acceptable model fit, without reaching the specified target in either study at all. Even worse, the good fitting ESEM solution could not be replicated across datasets. Only one factor (Factor 2) showed moderate evidence for replicability between studies. Upon further inspection, this factor consisted of four items targeting yearning for the deceased (Item 4), inability to trust others (Item 9), estrangement from others (Item 10) and avoiding to remember the deceased (Item 12). In our view, this observation should be interpreted with caution as content similarity between these statements seems to be rather low, and this factor was not present in either model identified in the psychometric review. In addition, we found that construct validity as measured by associations of ICG scores with depression and anxiety symptoms as measured with the PHQ-9 and GAD-7, respectively, did not depend on whether the ICG was modeled as a unidimensional construct or as a general factor with several group factors. This illustrates that the general factor not only captures most of the variance in ICG scores but also shows predictable associations with other constructs. Of course, the generalizability of this result depends on the choice of comparator instruments. Depression and anxiety were, however, most often used to investigate construct validity of the ICG and thus of high importance for the present study. Although the investigation of measurement properties of PHQ-9 and GAD-7 was beyond the scope of this study, we found that both instruments showed inadequate fit with a unidimensional CFA model. This is in line with recent research suggesting that both PHQ-9 and GAD-7 show inconsistent factorial validity (Boothroyd et al., 2019; Johnson et al., 2019; Maroufizadeh et al., 2019). Although further exploring the latent structure of these constructs poses a task for further research, PHQ-9 and GAD-7 sum scores (implicitly assuming unidimensional constructs) represent important clinical outcomes and are the basis for screening depression and anxiety disorders in primary care which is usually the point of entry

to care for people with complicated grief (Boothroyd et al., 2019; Costantini et al., 2021; Sapra et al., 2020). Therefore, we do not see this as a threat to our main conclusions. In general, our results regarding the ICG are in line with emerging evidence, that a vast number of psychological self-report measures fall prey to overfitting in validation studies (Rodriguez et al., 2016), which in turn impedes psychometric replication. Although some authors suggested their factor solutions to represent the latent nature of CG (e.g., Simon et al., 2011), we took a different perspective. We argued that to assess the degree to which apparent multidimensionality threatens the application of sum scores, one good fitting solution had to be found. We did not assume this solution to reflect latent psychological dimensions but rather unsystematic noise (cf. Cho et al., 2014). In fact, our results indicate that the ICG is essentially unidimensional, which is not to be equated with good model fit of a unidimensional CFA model (Reise, Bonifay, et al., 2013; Ziegler & Hagemann, 2015). With an increasing number of items, unidimensional CFA models will almost always lead to insufficient fit (Bentler, 2009). Taken together, (a) the degree of unsystematic variance inherent in the ICG items and (b) the large number of items to assess the unidimensional construct make it impossible to find a replicable measurement model for the ICG, while in practice sum scores offer a reliable estimate of CG.⁶ Consequently and given that there are no clear theoretical assumptions against which the latent structure of the ICG could be tested (except the unidimensional model proposed by Prigerson, Maciejewski, et al., 1995), structural equation modeling is not a well-suited method to evaluate the structural validity of the ICG as it will produce misleading and unreliable results. In fact, the inconsistencies regarding the factorial validity of the ICG might stem from validation studies overfitting unsystematic sources of variance using complex structural equation models, that are then interpreted as reflecting the latent nature of the construct of interest. One solution to resolve this problem might thus lie in reducing the number of ICG items based on clearly articulated theoretical grounds. For example, the ICG-derived Prolonged Grief Disorder-13 (PG-13) (Prigerson, Boelen, et al., 2021) consists of only 13 items to measure prolonged grief. However, there seems to be some indication that the PG-13 faces the same form of structural inconsistency as the ICG. For example, two recent validation studies of the PG-13 conclude that the PG-13 forms a unidimensional scale based on the eigenvalue criterion (Pohlkamp et al., 2018; Prigerson, Boelen, et al., 2021). Another study that applied CFA reports that the PG-13 forms a two- to three-dimensional construct (Sveen et al., 2020) mirroring the inconsistent results regarding the ICG. This is in line with some authors arguing that the eigenvalue criterion alone does not provide sufficient evidence for unidimensionality unless complemented with other methods of analysis and strong theoretical arguments (Ziegler & Hagemann, 2015).

In conclusion, the findings of the present study might very well also apply to measures derived from the ICG suggesting that not only the mere number of items of the ICG introduces unsystematic variance in item correlations, but also other factors such as the lack of theoretical scale development (i.e., item pool reduction for the ICG as well as the PG-13 was mainly based on empirical findings). Some limitations regarding the generalizability of these findings come to mind. As indicated earlier, validation studies of the ICG have relied on heterogeneous samples, which complicates evidence synthesis across studies. The present study relied on a specific population, namely bereaved caregivers of people with cancer in Germany which poses a potential constraint on generalizing our results. The samples in the present study were, however, similar to samples of other validation studies with respect to age, sex, relationship to the deceased, and mean ICG scores were neither extremely high nor extremely low in comparison to those reported in other studies (see Table A3 in Online Appendix A). As we held the study population constant between Sample 1 and Sample 2 and were still not able to replicate a measurement model to a satisfactory extent it is unlikely that population effects were the main drivers behind the structural inconsistencies reported in the literature. Still, sample effects (i.e., variation in samples from the same population affect the latent structure of the ICG) could explain our findings. Combining the findings from the psychometric review, SEM analyses and analyses on essential unidimensionality, we conclude that essential unidimensionality offers a more parsimonious and constructive explanation. That is, (a) if researchers needed to interpret the ICG differently in each sample (even if samples stem from the same population), then the generalizability of findings involving the ICG were severely constrained and (b) if inconsistent results regarding the factor structure of the ICG could be attributed to observable (or not observable) sample characteristics, then the ICG would be immune against threats on the structural validity and the statement that the instrument is structurally valid became tautological. Instead, essential unidimensionality allows to explain inconsistent factorial validity without immunizing the instrument against threats on structural validity and allows researchers to interpret ICG scores sensibly. Further research should, however, investigate potential sampling effects by controlling for sample characteristics using for instance post-stratification procedures (e.g., Deffner et al., 2021) or measurement invariance approaches (e.g., Marsh et al., 2020).

Conclusion

Implications for Clinical Research and Practice

Our results suggest that the ICG is essentially unidimensional and that previous reports of several subfacets are

misleading. For clinical researchers and clinicians alike, this finding has several important implications. First, evaluating associations of CG with other constructs of interest or investigating the extent of burden resulting from a loss is possible using ICG sum scores. Second, identifying associations of subfacets of CG with other constructs or investigating subfacet-specific CG profiles in a clinical population is neither necessary nor warranted. In fact, we would highly advise against the attempt to interpret potentially emerging subfacets of the ICG as distinct “symptom clusters” as other authors have suggested (Simon et al., 2011), because our results suggest that these clusters are not reliable and not replicable. Finally, if time constraints are an issue as is often the case in large survey studies or clinical contexts, then CG may just as well be measured by a reduced item set without a significant loss of information or complexity.

Implications for Psychometric Research

Our findings provide a parsimonious explanation for the seemingly paradox findings of the last 25 years of psychometric research on the ICG. On one hand, total scores reflect tantamount of variance in latent ICG scores and will produce predictable and reliable associations with other constructs such as depressive or anxiety symptoms. After all, it seems, that the factorial structure of the ICG is not as complicated as previous research on the ICG suggests. On the other hand, bad fit indices in CFA indicate a lack of unidimensionality and structural validity which undermines the theoretical conceptualization of the ICG. Hence, structural equation modeling is not well suited to evaluate the structural validity of the ICG in its current form. Further research should address these concerns through more theoretically inspired and parsimonious scale construction.

Acknowledgment

We would like to thank Ariane Preibsch, MSc, for assisting in the data collection.

Author Contributions (CRediT)

AS: Conceptualization, data curation, formal analysis, methodology, software, validation, visualization, writing—original draft, writing—review and editing; **JT:** Formal analysis, validation, writing—review and editing; **MH:** Conceptualization, funding acquisition, resources, supervision, writing—review and editing; **HCF:** Conceptualization, resources, supervision, writing—review and editing; **MWH:** Conceptualization, funding acquisition, resources, supervision, data curation, formal analysis, methodology, software, validation, writing—original draft, writing—review and editing.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this study was provided entirely by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) (Grant number HA 7536/1-1). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

ORCID iD

Markus W. Haun  <https://orcid.org/0000-0003-1851-3747>

Supplemental Material

Supplemental material for this article is available online.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Notes

1. We use the term to refer to any grief-related diagnosis, such as Prolonged Grief Disorder. There is some debate about the differences between CG and other grief-related conditions which to discuss lies beyond the scope of the present article. But see Maciejewski et al. (2016) and Prigerson, Kakarala, et al. (2021) for details. Note also that the items of the ICG do not perfectly align with diagnostic criteria for CG (Trembl et al., 2020).
2. It should be noted that most studies assessing criterion validity (3/4 studies) define the criterion (i.e., CG diagnosis) fully or partially based on ICG scores (e.g., Simon et al., 2011, define "caseness" as ICG scores above 30, clinical diagnosis of CG, self-diagnosis of CG, and loss at least 6 months in the past). Only Carmassi et al. (2014) defined the criterion independent of ICG scores.
3. Results converge on high correlations with depressive and anxiety symptoms ranging from $r = .38$ (Lumbeck et al., 2012) to $r = .75$ (Han et al., 2016) with 7/8 studies assessing construct validity reporting moderate to high correlations between CG and depressive symptoms/anxiety.
4. Note that Fisher et al. (2017) reported the same problem and chose to drop the sixth factor. We refrained from such a measure, because we saw no theoretical grounds to drop items and dropping items would make model comparison impossible.
5. Note, that in this case, model fit does not depend on the rotation method. Thus, the Bifactor ESEM 5+1-factor model fit is identical to that of the 6 factor ESEM model.

6. Note also that the reverse inference applies. As indicated by very low values of Ω_{H-S} reporting subscale scores can be highly misleading. Reise, Bonifay, et al. (2013) argue that under such conditions, the sum score of all items might be a better indicator of the subscales than the sum score of the subscale due to low reliability.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Author.
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16(3), 397–438. <https://doi.org/10.1080/10705510903008204>
- Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, 74(1), 137–143. <https://doi.org/10.1007/s11336-008-9100-1>
- Boothroyd, L., Dagnan, D., & Muncer, S. (2019). PHQ-9: One factor or two? *Psychiatry Research*, 271(December 2018), 532–534. <https://doi.org/10.1016/j.psychres.2018.12.048>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.
- Carmassi, C., Shear, M. K., Massimetti, G., Wall, M., Mauro, C., Gemignani, S., Conversano, C., & Dell'Osso, L. (2014). Validation of the Italian version Inventory of Complicated Grief (ICG): A study comparing CG patients versus bipolar disorder, PTSD and healthy controls. *Comprehensive Psychiatry*, 55(5), 1322–1329. <https://doi.org/10.1016/j.comppsy.2014.03.001>
- Cho, S.-J., Cohen, A. S., & Kim, S.-H. (2014). A mixture group bifactor model for binary responses. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 375–395. <https://doi.org/10.1080/10705511.2014.915371>
- Costantini, L., Pasquarella, C., Odone, A., Colucci, M. E., Costanza, A., Serafini, G., Aguglia, A., Belvederi Murri, M., Brakoulias, V., Amore, M., Ghaemi, S. N., & Amerio, A. (2021). Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): A systematic review. *Journal of Affective Disorders*, 279(March 2020), 473–483. <https://doi.org/10.1016/j.jad.2020.09.131>
- Deffner, D., Rohrer, J. M., & McElreath, R. (2021). A causal framework for cross-cultural generalizability. *PsyArXiv Preprints*. <http://arxiv.org/abs/2106.10576>
- Fisher, J. E., Mauro, C., Cozza, S. J., Wall, M., Simon, N. M., Ortiz, C. D., Harrington-LaMorie, J., Wang, Y., Fullerton, C. S., Ursano, R. J., & Katherine Shear, M. (2017). Examination of factor structure of the inventory of complicated grief (ICG) in a sample of bereaved military family members with persistent and elevated grief. *International Journal of Methods in Psychiatric Research*, 26, e1571–e1578. <https://doi.org/10.1002/mpr.1571>
- Han, D. H., Lee, J. J., Moon, D.-S., Cha, M.-J., Kim, M. A., Min, S., Yang, J. H., Lee, E. J., Yoo, S. K., & Chung, U.-S. (2016). Korean version of Inventory of Complicated Grief Scale: Psychometric properties in Korean adolescents. *Journal of Korean Medical Science*, 31(1), 114–119. <https://doi.org/10.3346/jkms.2016.31.1.114>
- Haun, M. W., Schakowski, A., Preibsch, A., Friederich, H. C., & Hartmann, M. (2019). Assessing decision regret in caregivers of deceased German people with cancer—A psychometric validation of the Decision Regret Scale for Caregivers.

- Health Expectations: An International Journal of Public Participation in Health Care and Health Policy*, 22(5), 1089–1099. <https://doi.org/10.1111/hex.12941>
- Hinz, A., Klein, A., & Brähler, E. (2017). Psychometric evaluation of the generalized anxiety disorder screener GAD-7, based on a large German general population sample. *Journal of Affective Disorders*, 210, 338–344.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huang, P. H. (2017). Asymptotics of AIC, BIC, and RMSEA for model selection in structural equation modeling. *Psychometrika*, 82(2), 407–426. <https://doi.org/10.1007/s11336-017-9572-y>
- Jacobs, S., Mazure, C., & Prigerson, H. (2000). Diagnostic criteria for traumatic grief. *Death Studies*, 24(3), 185–199. <https://doi.org/10.1080/074811800200531>
- Johnson, S. U., Ulvenes, P. G., Øktdalen, T., & Hoffart, A. (2019). Psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. *Frontiers in Psychology*, 10(JULY), 1–8. <https://doi.org/10.3389/fpsyg.2019.01713>
- Kroenke, K., Spitzer, R., & Williams, J. (2001). The PHQ-9. *General Internal Medicine*, 16(9), 606–613.
- Li, J., & Prigerson, H. (2016). Assessment and associated features of prolonged grief disorder among Chinese bereaved individuals. *Comprehensive Psychiatry*, 66, 9–16. <https://doi.org/10.1016/j.comppsy.2015.12.001>
- Lifshitz, R., Ifrah, K., Markovitz, N., Bluvstein, I., & Shmotkin, D. (2022). Inventory of complicated grief: Factor structure among middle-aged and older bereaved parents. *Death Studies*, 46, 200–207. <https://doi.org/10.1080/07481187.2020.1725931>
- Litalien, D., Morin, A. J. S., Gagné, M., Vallerand, R. J., Losier, G. F., & Ryan, R. M. (2017). Evidence of a continuum structure of academic self-determination: A two-study test using a bifactor-ESEM representation of academic motivation. *Contemporary Educational Psychology*, 51, 67–82. <https://doi.org/10.1016/j.cedpsych.2017.06.010>
- Ludwikowska-Świeboda, K., & Lachowska, B. (2019). Polish version of the Inventory of Complicated Grief—preliminary validation. *Psychiatria Polska*, 53(5), 1069–1086. <https://doi.org/10.12740/PP/91729>
- Lumbeck, G., Brandstätter, M., & Geissner, E. (2012). Erstvalidierung der deutschen version des “Inventory of Complicated Grief” (ICG-D) = Preliminary validation of the German version of the Inventory of Complicated Grief (ICG-D). *Zeitschrift Für Klinische Psychologie Und Psychotherapie: Forschung Und Praxis (Journal for Clinical Psychology and Psychotherapy: Research and Practice)*, 41(4), 243–248.
- Lundorff, M., Holmgren, H., Zachariae, R., Farver-Vestergaard, I., & O’Connor, M. (2017). Prevalence of prolonged grief disorder in adult bereavement: A systematic review and meta-analysis. *Journal of Affective Disorders*, 212(December 2016), 138–149. <https://doi.org/10.1016/j.jad.2017.01.030>
- Luo, Y., & Al-Harbi, K. (2016). The utility of the bifactor method for unidimensionality assessment when other methods disagree: An empirical illustration. *SAGE Open*, 6(4), 1–7. <https://doi.org/10.1177/2158244016674513>
- Maciejewski, P. K., Maercker, A., Boelen, P. A., & Prigerson, H. (2016). “Prolonged grief disorder” and “persistent complex bereavement disorder,” but not “complicated grief,” are one and the same diagnostic entity: An analysis of data from the Yale Bereavement Study. *World Psychiatry*, 15(3), 266–275. <https://doi.org/10.1002/wps.20348>
- Maroufizadeh, S., Omani-Samani, R., Almasi-Hashiani, A., Amini, P., & Sepidarkish, M. (2019). The reliability and validity of the Patient Health Questionnaire-9 (PHQ-9) and PHQ-2 in patients with infertility. *Reproductive Health*, 16(1), 4–11. <https://doi.org/10.1186/s12978-019-0802-x>
- Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (2020). Confirmatory Factor Analysis (CFA), Exploratory Structural Equation Modeling (ESEM), and Set-ESEM: Optimal balance between goodness of fit and parsimony. *Multivariate Behavioral Research*, 55(1), 102–119. <https://doi.org/10.1080/00273171.2019.1602503>
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the Big Five factor structure through Exploratory Structural Equation Modeling. *Psychological Assessment*, 22(3), 471–491. <https://doi.org/10.1037/a0019227>
- Masferrer, L., Garre-Olmo, J., & Caparrós, B. (2017). Estructura factorial y validez de constructo concurrente del IDC entre usuarios de drogas en duelo = Factor structure and concurrent construct validity of ICG among bereaved substance users. *Actas Españolas de Psiquiatria (Spanish Journal of Psychiatry)*, 45(2), 47–55.
- Mauro, C., Reynolds, C. F., Maercker, A., Skritskaya, N., Simon, N., Zisook, S., Lebowitz, B., Cozza, S. J., & Shear, M. K. (2019). Prolonged grief disorder: Clinical utility of ICD-11 diagnostic guidelines. *Psychological Medicine*, 49(5), 861–867. <https://doi.org/10.1017/S0033291718001563>
- Milic, J., Saavedra Perez, H., Zuurbier, L. A., Boelen, P. A., Rietjens, J. A., Hofman, A., & Tiemeier, H. (2019). The longitudinal and cross-sectional associations of grief and Complicated Grief with sleep quality in older adults. *Behavioral Sleep Medicine*, 17(1), 31–40. <https://doi.org/10.1080/15402002.2016.1276016>
- Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1171–1179. <https://doi.org/10.1007/s11136-017-1765-4>
- Mokkink, L. B., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2018). *COSMIN manual for systematic reviews of PROMs, user manual*. https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf
- Mokkink, L. B., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2019). *COSMIN study design checklist for patient-reported outcome measurement instruments* (Issue July). www.cosmin.nl
- Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, 23, 116–138.
- Morin, A. J. S., Arens, A. K., Tran, A., & Caci, H. (2016). Exploring sources of construct-relevant multidimensionality

- in psychiatric measurement: A tutorial and illustration using the composite scale of morningness. *International Journal of Methods in Psychiatric Research*, 25(4), 277–288. <https://doi.org/10.1002/mp>
- Morin, A. J. S., Meyer, J. P., Creusier, J., & Biétry, F. (2016). Multiple-group analysis of similarity in latent profile solutions. *Organizational Research Methods*, 19(2), 231–254. <https://doi.org/10.1177/1094428115621148>
- Muthén, B., & Muthén, L. (2019). *Mplus* (8.4th ed.). Muthén & Muthén.
- Niemand, T., & Mai, R. (2018). Flexible cutoff values for fit indices in the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 46(6), 1148–1172. <https://doi.org/10.1007/s11747-018-0602-9>
- Nye, C. D., & Drasgow, F. (2011). Assessing goodness of fit: Simple rules of thumb simply do not work. *Organizational Research Methods*, 14(3), 548–570. <https://doi.org/10.1177/1094428110368562>
- O'Connor, M., Lasgaard, M., Shevlin, M., & Guldin, M.-B. B. (2010). A confirmatory factor analysis of combined models of the Harvard Trauma Questionnaire and the Inventory of Complicated Grief-Revised: Are we measuring complicated grief or posttraumatic stress? *Journal of Anxiety Disorders*, 24(7), 672–679. <https://doi.org/10.1016/j.janxdis.2010.04.009>
- Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research and Evaluation*, 17(15), 1–8.
- Pohlkamp, L., Kreicbergs, U., Prigerson, H., & Sveen, J. (2018). Psychometric properties of the Prolonged Grief Disorder-13 (PG-13) in bereaved Swedish parents. *Psychiatry Research*, 267(January), 560–565. <https://doi.org/10.1016/j.psychres.2018.06.004>
- Prigerson, H., Boelen, P. A., Xu, J., Smith, K. V., & Maciejewski, P. K. (2021). Validation of the new DSM-5-TR criteria for prolonged grief disorder and the PG-13-Revised (PG-13-R) scale. *World Psychiatry*, 20(1), 96–106. <https://doi.org/https://doi.org/10.1002/wps.20823>
- Prigerson, H., Frank, E., Kasl, S. V., Reynolds, C. F., Anderson, B., Zubenko, G. S., Houck, P. R., George, C. J., & Kupfer, D. J. (1995). Complicated grief and bereavement-related depression as distinct disorders: preliminary empirical validation in elderly bereaved spouses. *The American Journal of Psychiatry*, 152(1), 22–30. <https://doi.org/10.1176/ajp.152.1.22>
- Prigerson, H., Kakarala, S., Gang, J., & Maciejewski, P. K. (2021). History and status of Prolonged Grief Disorder as a psychiatric diagnosis. *Annual Review of Clinical Psychology*, 17(1). <https://doi.org/10.1146/annurev-clinpsy-081219-093600>
- Prigerson, H., Maciejewski, P. K., Reynolds, C. F., Bierhals, A. J., Newsom, J. T., Fasiczka, A., Frank, E., Doman, J., & Miller, M. (1995). Inventory of Complicated Grief: A scale to measure maladaptive symptoms of loss. *Psychiatry Research*, 59(1–2), 65–79. [https://doi.org/10.1016/0165-1781\(95\)02757-2](https://doi.org/10.1016/0165-1781(95)02757-2)
- R Core Team. (2020). *R: A language and environment for statistical computing* (4.0.2). R Foundation for Statistical Computing. <https://www.r-project.org/>
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129–140. <https://doi.org/10.1080/00223891.2012.725437>
- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16, 19–31. <https://doi.org/10.1007/s11136-007-9183-7>
- Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in Structural Equation Modeling: A bifactor perspective. *Educational and Psychological Measurement*, 73(1), 5–26. <https://doi.org/10.1177/0013164412449831>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137–150. <https://doi.org/10.1037/met0000045>
- Sapra, A., Bhandari, P., Sharma, S., Chanpura, T., & Lopp, L. (2020). Using Generalized Anxiety Disorder-2 (GAD-2) and GAD-7 in a primary care setting. *Cureus*, 2(5). <https://doi.org/10.7759/cureus.8224>
- Shear, M. K. (2015). Complicated grief. *The New England Journal of Medicine*, 372(2), 153–160. <https://doi.org/10.1056/NEJMcp1315618>
- Shi, D., & Maydeu-Olivares, A. (2020). The effect of estimation methods on SEM fit indices. *Educational and Psychological Measurement*, 80(3), 421–445. <https://doi.org/10.1177/0013164419885164>
- Shi, D., Maydeu-Olivares, A., Rosseel, Y., Xia, Y., & Yang, Y. (2020). Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. *Structural Equation Modeling*, 27(1), 1–15. <https://doi.org/10.1080/10705511.2019.1611434>
- Simon, N. M., Shear, M. K., Reynolds, C. F., Cozza, S. J., Mauro, C., Zisook, S., Skritskaya, N., Robinaugh, D. J., Malgaroli, M., Spandorfer, J., & Lebowitz, B. (2020). Commentary on evidence in support of a grief-related condition as a DSM diagnosis. *Depression and Anxiety*, 37(1), 9–16. <https://doi.org/10.1002/da.22985>
- Simon, N. M., Wall, M. M., Keshaviah, A., Dryman, M. T., Leblanc, N. J., & Shear, M. K. (2011). Informing the symptom profile of complicated grief. *Depression and Anxiety*, 28(2), 118–126. <https://doi.org/10.1002/da.20775>
- Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, 66(4), 563–575. <https://doi.org/10.1007/bf02296196>
- Sveen, J., Bondjers, K., Heinsoo, J., & Arnberg, F. K. (2020). Psychometric evaluation of the Swedish version of the Prolonged Grief Disorder-13 (PG-13) in a bereaved mixed trauma sample. *Frontiers in Psychiatry*, 11, 1420.
- Terwee, C. B., Jansma, E. P., Riphagen, I. I., & De Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, 18(8), 1115–1123. <https://doi.org/10.1007/s11136-009-9528-5>
- Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W. J. G., Bouter, L. M., & De Vet, H. C. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN

- checklist. *Quality of Life Research*, 21(4), 651–657. <https://doi.org/10.1007/s11136-011-9960-1>
- Thimm, J., Davidsen, M., Eslness, M., & Vara, H. (2019). Reliability, factor structure, and validity of the Inventory of Complicated Grief (ICG) in a general bereaved sample in Norway. *Scandinavian Psychologist*, 6, e7. <https://doi.org/10.15714/scandpsychol.6.e7>
- Tönnies, J., Hartmann, M., Jäger, D., Bleyel, C., Becker, N., Friederich, H. -C., & Haun, M. W. (2021). *Aggressiveness of care at the end-of-life in cancer patients and its association with psychosocial functioning in bereaved caregivers*. *Frontiers in Oncology*, 11, 1895. <https://doi.org/10.3389/fonc.2021.673147>
- Treml, J., Kaiser, J., Plexnies, A., & Kersting, A. (2020). Assessing prolonged grief disorder: A systematic review of assessment instruments. *Journal of Affective Disorders*, 274(March), 420–434. <https://doi.org/10.1016/j.jad.2020.05.049>
- Wågø, S. S., Byrkjedal, I. K., Sinnes, H. M., Hystad, S. W., & Dyregrov, K. (2017). Social support and complicated grief: A longitudinal study on bereaved parents after the Utøya terror attack in Norway. *Scandinavian Psychologist*, 4, 1–22. <https://doi.org/10.15714/scandpsychol.4.e10>
- Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, 51(1), 409–428. <https://doi.org/10.3758/s13428-018-1055-2>
- Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items: Pitfalls and loopholes. *European Journal of Psychological Assessment*, 31(4), 231–237. <https://doi.org/10.1027/1015-5759/a000309>