Research Trends

Assessing the Reliability of the CAMS Rating Scale Using a Generalizability Study

Christopher D. Corona[1,2], Peter M. Gutierrez[3,4], Barry M. Wagner[5], and David A. Jobes[5]

[1]VISN 2 Center of Excellence for Suicide Prevention, Canandaigua VA Medical Center, Canandaigua, NY, USA
[2]Department of Psychiatry, University of Rochester Medical Center, NY, USA
[3]Rocky Mountain Mental Illness Research Education and Clinical Center, Denver Veterans Affairs Medical Center, Denver, CO, USA
[4]Department of Psychiatry, University of Colorado School of Medicine, Aurora, CO, USA
[5]Department of Psychology, Catholic University of America, Washington DC, USA

Abstract. *Background:* An important consideration when conducting randomized controlled trials is treatment differentiation. Direct observation helps ensure that providers in different treatment groups are delivering distinct interventions. One direct observation method is the use of a measure to rate clinician performance when delivering an intervention. *Aims:* This generalizability study evaluated the reliability of the CAMS Rating Scale (CRS), a measure used to assess delivery of the Collaborative Assessment and Management of Suicidality (CAMS). *Method:* Digitally recorded tapes of clinicians delivering either CAMS or Enhanced Care-As-Usual (E-CAU) were coded using the CRS. Sessions (*N* = 36) were each coded by two raters, and encompassed four clinicians, four time points, and 34 unique patients across two treatment groups. A reliability coefficient (i.e., G coefficient) and the percentages of variance contributed by each component of the measurement model were obtained. *Results:* The CRS reliably differentiates CAMS from E-CAU, minimizes measurement error relative to expected variance sources, and continues to demonstrate high inter-rater reliability. *Limitations:* The absence of blind raters, a formal training protocol for the rating team, and ratings from all clinician–patient dyads at all time points was a limitation. *Conclusion:* The CRS is a reliable treatment differentiation measure that can play an integral role in studies evaluating CAMS.

Keywords: suicide, intervention, research methods, fidelity, psychometrics

There were over 44,000 suicides in the United States in 2016, making it the 10th leading cause of death (Drapeau & McIntosh, 2017). Furthermore, suicide rates within the US military have more than doubled since 2004, surpassing the civilian rate for the first time in history (Bryan, Jennings, Jobes, & Bradley, 2012). These statistics highlight the need for the development and dissemination of effective clinical interventions for suicidal behavior in both military and civilian populations.

A white paper published by the American Psychological Association (APA) called for increased implementation of empirically supported psychotherapies, and specified the use of randomized controlled trials (RCTs) as an effective research design for identifying these interventions (APA, 1995). Furthermore, the need to demonstrate that experimental and control groups are receiving distinct treatments in RCTs underscores the assessment of fidelity as an important component of studies examining psychotherapeutic interventions (Shean, 2014).

Consensus regarding the definition of fidelity has not been reached; however, many definitions include treatment differentiation as an element (Bellg et al., 2004; Borrelli et al., 2005; Perepletchikova, Treat, & Kazdin, 2007). Without the assurance of distinct treatment groups, specific mechanisms responsible for differences in outcomes cannot be isolated. This can significantly hinder the ability to draw conclusions about the effects of different interventions on outcomes (Kazdin, 2003; Perepletchikova et al., 2007).

Despite the importance of treatment differentiation and other aspects of fidelity, there is little evidence that researchers are adequately incorporating these constructs into their studies (Perepletchikova et al., 2007; Prowse, Nagel, Meadows, & Enticott, 2015). However, Smith, Daunic, and Taylor (2007) argue that methods for incorporating fidelity procedures into treatment studies do exist. Furthermore, the Treatment Fidelity Workgroup of the National Institutes of Health Behavior Change Consortium stipulates that one method is the monitoring of intervention delivery throughout a study (Bellg et al., 2004).

Lane, Bocian, MacMillan, and Gresham (2004) outlined methods for monitoring and assessing treatment delivery, including direct observation, consultant feedback, self-monitoring, and treatment manualization. While

rigorously designed studies ideally incorporate several of these elements, direct observation is the most highly recommended (Lane et al., 2004; Smith et al., 2007). An extensively used methodology for direct observation is the development of a measure that can be used to rate clinician performance regarding specific components of a particular intervention.

Cognitive therapy (CT) is currently the only empirically supported treatment for suicidal behavior with a corresponding measure that has been extensively used and psychometrically studied (Barber, Liese, & Abrams, 2003; Crits-Christoph et al., 1998; Dobson, Shaw, & Vallis, 1985; Shaw et al., 1999; Vallis, Shaw, & Dobson, 1986; Whisman, 1993). Some have employed this measure in trials assessing the effect of CT on suicidal behavior (Brown et al., 2005; Rudd et al., 2015); however, specific analyses of its psychometric properties have not been conducted using data from these trials. While a DBT Global Rating Scale has been developed (Linehan, 2003), it is not discussed extensively in the literature and has not been psychometrically evaluated thus far in published research. Considering both the urgent need for empirically supported treatments aimed at reducing suicidal behavior and the key role that treatment differentiation plays in randomized controlled trials, there is considerable need for psychometrically sound measures that can be used to demonstrate this component of fidelity within studies of suicide-specific interventions.

The Collaborative Assessment and Management of Suicidality (CAMS; Jobes, 2016; Jobes, Comtois, Brenner, Gutierrez, & O'Connor, 2016) is a short-term therapeutic framework that has demonstrated effectiveness in three previous RCTs (Andreasson et al., 2016; Comtois et al., 2011; Jobes et al., 2017). Designed to be employed by clinicians from different training backgrounds, CAMS encompasses various tasks imperative to the treatment of suicidal behavior including risk assessment, treatment planning, and interventions that target patient-defined suicidal *drivers*. In addition to focusing on issues relevant to suicide, a collaborative style is one of the hallmarks of CAMS that frames all aspects of care delivery (Jobes, 2016).

The CAMS Rating Scale (CRS) is a measure that was developed to assess adherence to CAMS (Comtois et al., 2011; Jobes, 2012). The measure incorporates items that are designed to capture the overarching style of CAMS and technical components that are used at different points throughout care. These items are organized into domains entitled *Collaboration, Suicide Focus, Risk Assessment, Treatment Planning,* and *Intervention*. An additional item measuring general adherence to the framework is also included. While certain subscales (i.e., Risk Assessment, Treatment Planning, and Intervention) were designed to partial out distinct components of CAMS care, other subscales (i.e.,

Collaboration and Suicide Focus) were designed to capture a style that should be present across all components of the framework. A previous analysis of the CRS demonstrated high interrater reliability using intraclass correlation coefficients (ICCs; Corona & Jobes, 2015). However, ICCs are calculated using only one source of variance at a time, and do not simultaneously account for other sources of variance that could affect reliability in complex measurement models (Mushquash & O'Connor, 2006; Wasserman, Levy, & Loken, 2009).

Generalizability (G) studies are designed to assess precision in measurement models that incorporate multiple sources of variance (Cardinet, Johnson, & Pini, 2010). While accurate measurement is important to any study, the need to demonstrate precision is especially relevant to psychotherapy trials characterized by numerous sources of measurement variance (Wasserman et al., 2009). More specifically, G studies can provide information about how reliably a treatment measure differentiates within a particular measurement component while accounting for other sources of variance (Cardinet et al., 2010). For example, a G study can indicate how reliably a measure differentiates between treatment groups while accounting for differences among raters, clinicians, sessions, and test items. G studies also provide estimates of the variance contributed by each measurement component, and researchers can then determine whether this variance is coming from expected or unwanted sources (Mushquash & O'Connor, 2006). These properties highlight the utility of G studies in demonstrating the ability of a measure to assess treatment differentiation.

The goal of the current G study was to evaluate the ability of the CRS to reliably differentiate between CAMS and another treatment, and to determine the contributions of variance from different sources in the measurement model. This is the first known study that applies this methodology to a measure assessing a suicide-specific intervention. The primary hypothesis of the current study posited that the CRS would reliably differentiate CAMS from another treatment. A secondary hypothesis predicted that variance in the measurement model would come primarily from clinicians and CRS items. The rationale for this prediction is that there is inherent variability in the practice of different clinicians, especially those delivering different interventions. Furthermore, clinicians delivering CAMS were expected to elicit different scores on items assessing adherence to CAMS than control clinicians. A tertiary hypothesis predicted that raters would account for minimal variance in CRS ratings given that the measure has previously demonstrated high inter-rater reliability (Corona & Jobes, 2015).

Method

The current IRB-approved study used adherence ratings of digitally recorded therapy sessions that were collected as part of the Operation Worth Living (OWL) study, an RCT comparing CAMS with Enhanced Care-As-Usual (E-CAU) conducted at an army post in the southeastern United States. In the OWL study, treatment duration in the CAMS group ranged from one to 26 sessions, with a median of five ($M = 6.2$, $SD = 3.9$). In the E-CAU group, treatment ranged from one to 21 treatment sessions, also with a median of five ($M = 6.4$, $SD = 3.5$). For further information regarding the methodology and sample characteristics of the OWL study, refer to Jobes et al. (2017).

Patients in the current study ($N = 34$) included adult, active duty army soldiers identified as suicidal based on Beck Scale for Suicide Ideation scores (BSS; Beck, Steer, & Ranieri, 1988). BSS scores at or above 6 indicate clinically significant suicide ideation, and baseline scores in the current sample ranged from 13 to 31 ($M = 20.0$; $SD = 5.0$). The number of lifetime suicide attempts in the sample per self-report ranged from 0 to 222 ($M = 10.9$; $SD = 41.0$). However, 50% of the sample reported zero lifetime suicide attempts, and 88.2% of the sample reported between zero and four lifetime suicide attempts. Furthermore, the modal number of suicide attempts was 0. The sample ranged in age from 19 to 48 years ($M = 27.6$; $SD = 6.8$). Those identifying as male comprised 88.2% of the sample, and 11.8% identified as female. Enlisted army soldiers comprised 94.1%, and officers comprised 5.9% of the sample.

In all, 19 patients in the current study (52.8%) were assigned to the E-CAU condition, while 17 (47.2%) were assigned to the CAMS condition. Clinicians included licensed clinical social workers ($N = 4$) delivering CAMS ($N = 2$) or E-CAU ($N = 2$). Providers in the CAMS group were trained in the therapeutic framework, which included being provided with copies of the manual, in-person training, and participation in case conferences for clinical consultation. E-CAU clinicians did not receive training in CAMS, and responded to suicide-related clinical issues in a manner that was consistent with their training. Because CAU was provided within the context of an RCT, it was enhanced by close monitoring and follow-up related to session attendance and the standardized administration of multiple assessments throughout the study.

The CRS consists of 14 total items, each of which is rated on a 7-point Likert scale ranging from 0 (*poor*) to 6 (*excellent*), with 3 representing *satisfactory* performance. The first 11 items are organized into domains that capture key components of the CAMS framework. These domains are Collaboration, Suicide Focus, Risk Assessment, Treatment Planning, and Intervention. Item 12 assesses overall adherence to the framework. Items 13 and 14 encompass general elements of the therapeutic process, do not represent essential components of the framework, and are thus

Table 1. CAMS Rating Scale (CRS) items

Item	Description
Collaboration	The clinician expressed empathy with the patient's suicidal wish. All assessments were conducted interactively with substantial input from both clinician and patient. The treatment plan was designed and modified interactively with substantial input from both clinician and patient. All interventions (in-session) were selected and modified interactively with substantial input and participation from both clinician and patient.
Suicide focus	The clinician clarified the CAMS agenda to focus on factors related to the suicidal wish any time it appeared needed; when factors arose that were not directly or indirectly leading to suicidality for this patient, they were acknowledged as important, but not the focus of the current work.
Risk assessment	The dyad followed the framework for initiating and completing the SSF assessment at the beginning of the session.
Treatment planning	The dyad developed and/or updated a Crisis Response Plan (CRP), Safety Plan (SP), or CAMS Stabilization Plan (CSP), which includes regularly attending therapy sessions, addressing barriers to care, means restriction, decreasing isolation, and use of a coping card. The treatment plan identified and targeted the most relevant direct and/or indirect drivers of suicidality as determined by the dyad. The treatment plan established the use of suicide-specific, problem-focused interventions to target and treat the drivers of the patient's suicidality.
Intervention	The session included the use of suicide-specific, problem-focused interventions to target and treat the drivers of suicidality. The session included a discussion about hope, reasons for living, plans, goals, purpose, and meaning.
Overall adherence	How would you rate the clinician's overall adherence to the CAMS framework?
General items	How receptive was the patient to this model of treatment? How comfortable did the clinician seem?

Note. CAMS = Collaborative Assessment and Management of Suicidality.

not included in adherence assessments. All CRS items are listed in Table 1.

Initially, each clinician was assessed to ensure that delivery of treatment was consistent with their assigned treatment group. A team of two CAMS experts watched and rated initial sessions until criteria for adherence to respective treatment groups were met. Clinicians were considered initially adherent to CAMS after four consecutive sessions that met the following criteria: an average score of 3 or higher in each CRS domain and an average score of 3 or higher on Item 12. Scores for each CRS domain were obtained by averaging ratings across raters and items in that domain. Clinicians in the E-CAU group were considered initially nonadherent after four consecutive sessions that did not meet all CAMS adherence criteria.

Following initial adherence assessments, a team of two raters (i.e., one CAMS expert and one doctoral student with extensive CAMS training) rated (i.e., spot-checked) every 10th session from each clinician in each treatment group to ensure fidelity throughout the study. The same criteria for adherence and nonadherence were applied to spot-check sessions. Any fidelity violations were addressed via consultation between study personnel and clinicians. In all cases, raters were not blind to treatment groups.

The current study included 36 sessions that were all rated by the same two raters using the CRS. These sessions encompassed four different clinicians (two from each treatment group), four different sessions throughout the course of treatment (i.e., Sessions 1–4), and 34 different patients. Table 2 displays the number of sessions at different time points in treatment (i.e., Sessions 1–4) included for each clinician. Values greater than 1 indicate that more than one patient was seen by the clinician at that time point. In these instances, item-level scores were averaged across patients to create one set of scores for each clinician at each time point.

Analyses in the current study were conducted using EduG version 6.1 (Cardinet et al., 2010). Sources of measurement variance defined in the analysis were CRS items, raters, treatment groups, clinicians, and sessions. Treatment group was designated as the differentiating factor, which allowed for the evaluation of whether the CRS reliably differentiates between CAMS and E-CAU in the current sample.

Statistics interpreted from this analysis included the percentages of variance contributed by each source (including all interactions) and the absolute G coefficient (Coef_G). When treatment group is designated as the differentiating factor, this coefficient is a ratio of treatment group variance to total variance (i.e., the combination of treatment group variance and error variance) when accounting for other sources of variance in the measurement model (i.e., CRS items, raters, clinicians, and sessions; Cardinet et al., 2010).

Results and Discussion

The absolute G coefficient (Coef_G) for this measurement model was 0.98. Furthermore, the three sources contributing the most to variance were clinicians (25.4%), the interaction between treatment groups and CRS items (20.4%), and the interaction between clinicians and sessions (17%). Notably, raters contributed less than 0.1% to the total variance. The contributions of variance from all sources (including all interactions) are presented in Table 3.

The goal of the current study was to assess the ability of the CRS to reliably function as a treatment differentiation measure using a G study. Treatment group was thus used as the differentiating factor in this analysis. Moreover, simultaneously assessing the variance accounted for by other components of the measurement model provided additional indicators of the reliability of the measure. It was predicted that the CRS would be able to reliably differentiate CAMS from another treatment (i.e., E-CAU). The high reliability coefficient (Coef_G) obtained in the current study indicates that measurement error is small relative to between-treatment group variance when accounting for other sources of variance (i.e., clinicians, raters, test items, and sessions). This finding supports the initial prediction, and suggests that the CRS functions well as a treatment differentiation measure.

Table 2. Number of rated sessions

	CAMS		E-CAU		
	Clinician 1	Clinician 2	Clinician 3	Clinician 4	Total
Session 1	3	2	1	4	10
Session 2	2	1	2	2	7
Session 3	2	2	2	3	9
Session 4	4	1	2	3	10
Total	11	6	7	12	36

Note. The same two raters coded all sessions, which encompassed 34 different patients. CAMS = Collaborative Assessment and Management of Suicidality. E-CAU = Enhanced Care-As-Usual.

Table 3. Percentage of variance contributed by measurement components (Coef_G = 0.98)

Source	Differentiation variance	Absolute variance	% Absolute variance
G	3.154		
C		0.01504	25.4
GxI		0.01203	20.4
CxS		0.01005	17
CxR		0.00586	9.9
GxRxI		0.0034	5.8
I		0.0027	4.6
RxI		0.00208	3.5
RxS		0.0019	3.2
CxI		0.00156	2.6
CxSxI		0.00129	2.2
GxRxS		0.00124	2.1
CxRxSxI		0.00098	1.7
CxRxS		0.00037	0.6
GxRxS		0.00024	0.4
RxSxI		0.00018	0.3
CxRxI		0.00012	0.2
GxSxI		< 0.001	0.1
R		< 0.001	< 0.1
S		< 0.001	< 0.1
GxR		< 0.001	< 0.1
GxS		< 0.001	< 0.1
SxI		< 0.001	< 0.1
Total	3.154	0.05912	100

Note. G = treatment groups; C = clinicians; I = items; R = raters; S = sessions; interactions among components are denoted with an *x*.

It was also predicted that variance in the measurement model would come primarily from clinicians and CRS items because there is inherent variability in the practice of different clinicians, especially those delivering different interventions. Furthermore, clinicians delivering CAMS were expected to elicit different scores on items assessing adherence to CAMS than clinicians delivering E-CAU. Results support these predictions, with most variance coming from clinicians (25.4%) and the interaction between CRS items and treatment groups (20.4%). The latter finding suggests that, as expected, CRS scores were different for clinicians in different treatment groups.

Another notable source of variance was the interaction between clinicians and sessions (17%). Although not hypothesized, this finding suggests that performance varied by session among clinicians in the current study. While this is generally expected in the delivery of psychotherapeutic interventions, it is especially so with regard to CAMS given that this intervention emphasizes different elements depending on the session. Early CAMS sessions are characterized by risk assessment and treatment planning (i.e., identification of problems driving the desire for suicide), whereas later sessions are characterized by targeted intervention and resolution (i.e., debriefing the treatment process). Thus, this is an expected source of variance in scores measuring adherence to CAMS as a whole because different elements of the intervention are expected to be emphasized in different sessions.

Also predicted was that raters would account for some variance, although the magnitude was expected to be low given that the CRS has previously demonstrated high inter-rater reliability (Corona & Jobes, 2015). Notably, less than 0.1% of the variance was accounted for by raters, suggesting that different raters who independently rated treatment sessions contributed minimal variance to measurement. These findings provide further support for the interrater reliability of the CRS.

Limitations

The results presented should be considered within the context of certain limitations, which included the absence both of blind raters and of a formal training protocol for the rating team. Owing to the need to adhere to strict methodological requirements (i.e., regularly spot-checking sessions from both treatment groups) with limited study resources, only a small team of raters was approved for the study. This team was thus responsible for watching and rating videos from both treatment groups. Furthermore, the rating team comprised personnel who contributed significantly to the development of CAMS and were considered experts with regard to the framework. Thus, only minimal training was provided to raters in the current study. Future studies should employ a formalized training process for raters that can be replicated in other settings.

Another notable limitation pertains to the methodology of the OWL study. In an earlier RCT comparing CAMS with E-CAU (Comtois et al., 2011), patients receiving CAMS were seen for an average of 8.4 sessions. Accordingly, the spot-check protocol in the OWL study stipulated that clinicians would be rated on every 10th session they delivered. This was done to minimize the likelihood of repeatedly rating the same clinician–patient dyad, thus obtaining a more general indication of clinicians' performance. As a result, sets of ratings for clinicians in the current study often include more than one of a particular session number (e.g., multiple first sessions), but with different patients. Because ratings were not collected from all clinician–patient dyads at all time points, we averaged ratings for each

clinician from a given session number across all patients seen at that time point. Future G studies would ideally incorporate a design that rates all clinician–patient dyads at all time points.

Also worth discussing is the difficulty inherent in ensuring that adherence raters remain truly blind in any RCT involving specialized psychotherapeutic interventions. Standard practice entails raters having some training in the intervention they could be coding in the interest of providing valid assessments of adherence to that intervention, as coders unaware of what they should be looking for stand to threaten the validity of such ratings (Carroll et al., 2000). Furthermore, one could subsequently argue that using even minimally trained coders when studying a specialized intervention precludes the possibility of truly blind ratings regardless of whether they are explicitly told beforehand what they are observing. This is because any intervention with clear hallmarks (e.g., in CAMS, the use of the Suicide Status Form and assessment of the specific constructs contained therein, or the discussion of suicidal drivers during session) could provide clues for coders as to the intervention being delivered and ultimately break any blinding. However, we believe this should not preclude the evaluation of adherence measures for specialized interventions, as they represent an important component of intervention development and dissemination.

Conclusion

The results of this G study underscore important characteristics of the CRS: its ability to reliably differentiate CAMS from another treatment (i.e., E-CAU), its ability to minimize measurement error relative to expected sources of variance (i.e., from different clinicians, test items, and sessions), and its continued demonstration of high interrater reliability. These properties establish the CRS as a measure that can play an integral role in RCTs evaluating CAMS. Furthermore, this G study contributes to the literature on methods for developing treatment differentiation measures, which is important as the field of suicidology continues to evaluate clinical interventions using RCTs.

Acknowledgments

References

American Psychological Association, Task Force on Psychological Intervention Guidelines. (1995). *Template for developing guidelines: Interventions for mental disorders and psychosocial aspects of physical disorders.* Washington, DC: American Psychological Association.

Andreasson, K., Krogh, J., Wenneberg, C., Jessen, H. K., Krakauer, K., Gluud, C., ... Nordentoft, M. (2016). Effectiveness of dialectical behavior therapy versus collaborative assessment and management of suicidality treatment for reduction of self-harm in adults with borderline personality traits and disorder – A randomized observer-blinded clinical trial. *Depression and Anxiety, 33*(6), 520–530. https://doi.org/10.1002/da.22472

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. https://doi.org/10.1093/ptr/kpg019

Beck, A. T., Steer, R. A., & Ranieri, W. F. (1988). Scale for suicide ideation: Psychometric properties of a self-report version. *Journal of Clinical Psychology, 44*(4), 499–505.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., ... Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology, 23*(5), 443–451. https://doi.org/10.1037/0278-6133.23.5.443

Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., ... Orwig, D. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology, 73*(5), 852–860. https://doi.org/10.1037/0022-006X.73.5.852

Brown, G. K., Have, T. T., Henriques, G. R., Xie, S. X., Hollander, J. E., & Beck, A. T. (2005). Cognitive therapy for the prevention of suicide attempts: A randomized controlled trial. *Journal of the American Medical Association, 294*(5), 563–570. https://doi.org/10.1001/jama.294.5.563

Bryan, C. J., Jennings, K. W., Jobes, D. A., & Bradley, J. C. (2012). Understanding and preventing military suicide. *Archives of Suicide Research, 16*(2), 95–110. https://doi.org/10.1080/13811118.2012.667321

Cardinet, J., Johnson, J., & Pini, G. (2010). *Applying generalizability theory using EduG.* New York, NY: Routledge. Retrieved from https://doi.org/10.4324/9780203866948

Carroll, K. M., Nich, C., Sifry, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., ... Rounsaville, B. J. (2000). A general system for evaluating therapist adherence and competence in psychotherapy re-

search in the addictions. *Drug & Alcohol Dependence, 57*, 225–238. https://doi.org/10.1016/S0376-8716(99)00049-6

Comtois, K. A., Jobes, D. A., S. O'Connor, S., Atkins, D. C., Janis, K., E. Chessen, C., … Yuodelis-Flores, C. (2011). Collaborative assessment and management of suicidality (CAMS): Feasibility trial for next-day appointment services. *Depression and Anxiety, 28*(11), 963–972. https://doi.org/10.1002/da.20895

Corona, C. D., & Jobes, D. A. (2015, April). *The psychometric properties of the CAMS rating scale: A preliminary evaluation*. Paper session presented at the 48th Annual Conference of the American Association of Suicidology, Atlanta, GA.

Crits-Christoph, P., Siqueland, L., Chittams, J., Barber, J. P., Beck, A. T., Frank, A., … Woody, G. (1998). Training in cognitive, supportive-expressive, and drug counseling therapies for cocaine dependence. *Journal of Consulting and Clinical Psychology, 66*(3), 484–492. https://doi.org/10.1037//0022-006x.66.3.484

Dobson, K. S., Shaw, B. F., & Vallis, T. M. (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology, 24*(4), 295–300. https://doi.org/10.1111/j.2044-8260.1985.tb00662.x

Drapeau, C. W., & McIntosh, J. L. (2017). *U.S.A. suicide 2016: Official final data*. American Association of Suicidology: Washington, DC. Retrieved from http://www.suicidology.org/resources/facts-statistics

Jobes, D. A. (2012). The collaborative assessment and management of suicidality (CAMS): An evolving evidence-based clinical approach to suicidal risk. *Suicide and Life-Threatening Behavior, 42*(6), 640–653. https://doi.org/10.1111/j.1943-278x.2012.00119.x

Jobes, D. A. (2016). *Managing suicidal risk: A collaborative approach* (2nd ed.). New York, NY: Guilford Press.

Jobes, D., Comtois, K., Brenner, L., Gutierrez, P., & O'Connor, S. (2016). Lessons learned from clinical trials of the collaborative assessment and management of suicidality (CAMS). In R. O'Connor, S. Platt, & J. Gordon (Eds.), *International handbook of suicide prevention: Research, policy, & practice* (2nd ed., pp. 383–400). West Sussex, UK: Wiley-Blackwell. Retrieved from https://doi.org/10.1002/9781118903223.ch24

Jobes, D. A., Comtois, K A., Gutierrez, P. M., Brenner, L., Huh, D., Chalker, S., … Crow, B. (2017). A randomized controlled trial of the collaborative assessment and management of suicidality versus enhanced care as usual with suicidal soldiers. *Psychiatry: Interpersonal and Biological Processes, 80*(4), 339–356. https://doi.org/10.1080/00332747.2017.1354607

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential – but often forgotten – component of school-based interventions. *Preventing School Failure: Alternative Education for Children and Youth, 48*(3), 36–43. https://doi.org/10.3200/psfl.48.3.36-43

Linehan, M. M. (2003). *DBT global rating scale*. Unpublished manuscript.

Mushquash, C., & O'Connor, B. P. (2006). SPSS and SAS programs for generalizability theory analyses. *Behavior Research Methods, 38*(3), 542–547. https://doi.org/10.3758/bf03192810

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. https://doi.org/10.1037/0022-006x.75.6.829

Prowse, P. D., Nagel, T., Meadows, G. N., & Enticott, J. C. (2015). Treatment fidelity over the last decade in psychosocial clinical trials outcome studies: A systematic review. *Journal of Psychiatry, 18*, 258. https://doi.org/10.4172/psychiatry.1000258

Rudd, M. D., Bryan, C. J., Wertenberger, E. G., Peterson, A. L., Young-McCaughan, S., Mintz, J., … Delano, K. (2015). Brief cognitive-behavioral therapy effects on post-treatment suicide attempts in a military sample: Results of a randomized clinical trial with 2-year follow-up. *American Journal of Psychiatry, 172*(5), 441–449. https://doi.org/10.1176/appi.ajp.2014.14070843

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. https://doi.org/10.1037//0022-006x.67.6.837

Shean, G. (2014). Limitations of randomized control designs in psychotherapy research. *Advances in Psychiatry, 2014*, 561452. https://doi.org/10.1155/2014/561452

Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education and Treatment of Children, 30*(4), 121–134. https://doi.org/10.1353/etc.2007.0033

Vallis, T. M., Shaw, B. F., & Dobson, K. S. (1986). The cognitive therapy scale: Psychometric properties. *Journal of Consulting and Clinical Psychology, 54*(3), 381–385. https://doi.org/10.1037//0022-006x.54.3.381

Wasserman, R. H., Levy, K. N., & Loken, E. (2009). Generalizability theory in psychotherapy research: The impact of multiple sources of variance on the dependability of psychotherapy process ratings. *Psychotherapy Research, 19*(4-5), 397–408. https://doi.org/10.1080/10503300802579156

Whisman, M. A. (1993). Mediators and moderators of change in cognitive therapy of depression. *Psychological Bulletin, 114*(2), 248–265. https://doi.org/10.1037//0033-2909.114.2.248

Christopher D. Corona, PhD, is a postdoctoral fellow with appointments in the VISN 2 Center of Excellence for Suicide Prevention at the Canandaigua VA Medical Center as well as the Department of Psychiatry at the University of Rochester Medical Center, NY.

Peter M. Gutierrez, PhD, is a clinical research psychologist with appointments in the Rocky Mountain Mental Illness Research Education and Clinical Center at the Denver VA Medical Center as well as the Department of Psychiatry at the University of Colorado School of Medicine, CO.

Barry M. Wagner, PhD, is a professor in the Department of Psychology at the Catholic University of America, Washington, DC. His research focuses on identifying family relational factors and adolescent emotion regulation styles that influence ongoing suicide risk.

David A. Jobes, PhD, is a professor and Associate Director of Clinical Training in the Department of Psychology at the Catholic University of America, Washington, DC. His research is centered on clinical risk assessment (using both quantitative and qualitative methods) with different suicidal populations in different clinical settings.

Christopher D. Corona
VISN 2 Center of Excellence for Suicide Prevention
Canandaigua VAMC
400 Fort Hill Avenue
Canandaigua, NY 14424
USA
christopher.corona2@va.gov