



Optimized short-forms of the Cognitive Distortions Questionnaire

Amanda S. Morrison^{a,*}, Berk Ustun^b, Arielle Horenstein^c, Simona C. Kaplan^c,
 Irismar Reis de Oliveira^d, Sedat Batmaz^e, James J. Gross^f, Ekaterina Sadikova^{g,h},
 Curt Hemannyⁱ, Pedro P. Pires^j, Philippe R. Goldin^k, Ronald C. Kessler^h, Richard G. Heimberg^c

^a Department of Psychology, California State University, East Bay, Hayward, CA, USA

^b Halıcıoğlu Data Science Institute, UC San Diego, La Jolla, CA, USA

^c Department of Psychology and Neuroscience, Temple University, Philadelphia, PA, USA

^d Department of Neurosciences and Mental Health, Federal University of Bahia, Bahia, Brazil

^e Department of Psychology, Faculty of Social Sciences and Humanities, Social Sciences University of Ankara, Ankara, Turkey

^f Department of Psychology, Stanford University, Stanford, CA, USA

^g Department of Epidemiology, Harvard Chan School of Public Health, Boston, MA, USA

^h Department of Health Care Policy, Harvard Medical School, Boston, MA, USA

ⁱ Postgraduate Program of Interactive Processes of Organs and Systems, Health Sciences Institute, Federal University of Bahia, Bahia, Brazil

^j Department of Psychometrics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

^k Betty Irene Moore School of Nursing, University of California, Davis, Davis, CA, USA

ARTICLE INFO

Keywords:

Cognitive distortions
 Cognitive errors
 Cognitive behavioral therapy
 Assessment
 Short form

ABSTRACT

Introduction: The Cognitive Distortions Questionnaire (CD-Quest) is a self-report questionnaire that assesses common cognitive distortions. Although the CD-Quest has excellent psychometric properties, its length may limit its use.

Methods: We attempted to develop short-forms of the CD-Quest using RiskSLIM – a machine learning method to build short-form scales that can be scored by hand. Each short-form was fit to maximize concordance with the total CD-Quest score for a specified number of items based on an objective function, in this case R^2 , by selecting an optimal subset of items and an optimal set of small integer weights. The models were trained in a sample of US undergraduate students ($N = 906$). We then validated each short-form on five independent samples: two samples of undergraduate students in Brazil ($Ns = 182, 183$); patients with depression in Brazil ($N = 62$); patients with social anxiety disorder in the US ($N = 198$); and psychiatric outpatients in Turkey ($N = 269$).

Results: A 9-item short-form with integer scoring was created that reproduced the total 15-item CD-Quest score in all validation samples with excellent accuracy ($R^2 = 90.4\text{--}93.6\%$). A 5-item ultra-short-form had good accuracy ($R^2 = 78.2\text{--}85.5\%$).

Discussion: A 9-item short-form and a 5-item ultra-short-form of the CD-Quest both reproduced full CD-Quest scores with excellent to good accuracy. These shorter versions of the full CD-Quest could facilitate measurement of cognitive distortions for users with limited time and resources.

1. Introduction

Cognitive distortions are erroneous, irrational, or exaggerated ways of thinking that are thought to play a key role in the development and persistence of many mental disorders. In Beck's (1976) cognitive theory of depression, which informs modern cognitive behavioral therapy (CBT), cognitive distortions are seen as stemming from rigidly-held beliefs about the self, world, and future. As a result of these cognitive distortions, some individuals are predisposed to experience negative

automatic thoughts, which subsequently induce pathological emotional and behavioral outcomes. Beck's therapy teaches individuals to identify their cognitive distortions and challenge their negative automatic thoughts.

Several self-report scales have been developed to assess overall burden of cognitive distortions, including the Cognitive Error Questionnaire – General Form (CEQ; Lefebvre, 1981), the Inventory of Cognitive Distortions (ICD; Yurica, 2002), the Cognitive Distortions Scale (CDS; Covin et al., 2011), and the Cognitive Distortions

* Correspondence to: Department of Psychology, California State University, East Bay, 25800 Carlos Bee Boulevard, Hayward, CA 94542, USA.

E-mail address: amanda.morrison@csueastbay.edu (A.S. Morrison).

<https://doi.org/10.1016/j.janxdis.2022.102624>

Received 10 July 2021; Received in revised form 13 August 2022; Accepted 17 August 2022

Available online 20 August 2022

0887-6185/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Questionnaire (CD-Quest; de Oliveira et al., 2015a). Of these scales, the one with the strongest psychometric evidence is the CD-Quest. The CD-Quest assesses frequency and intensity of 15 common cognitive distortions. The scale has been found to exhibit very good psychometric properties in both undergraduate student samples (de Oliveira et al., 2015; Kostoglou & Pidgeon, 2016; Morrison et al., 2015; Qian et al., 2020) and clinical samples (Batmaz et al., 2015; Kaplan et al., 2017) as well as to have good sensitivity to change and predictive validity (Butler et al., 2021; Hemanny et al., 2019). These results have led to the CD-Quest being widely used in clinical settings for measurement-based care.

The CD-Quest is a complex and time-consuming scale in that it begins by presenting users with an orientating introductory statement of over 300 words and then asks users to make ratings of 15 separate statements regarding their cognitive distortions in a 9-cell cross-classification of the extent to which the statement is believed (3 categories of *a little [up to 30%]*, *much [31% to 70%]*, *very much [more than 70%]*) by the frequency of the distortion's occurrence (3 categories of *occasionally [1-2 days during the past week]*, *much of the time [3-5 days during the past week]*, *almost all of the time [6-7 days during the past week]*). In addition to the complexity of the rating task itself, ambiguity is introduced by the use of double-barreled response options (e.g., 5 minutes every day over the past week would not be *almost all of the time* but it would be *6-7 days during the past week*). Despite these complexities, though, the methodological studies cited above show that the scale has demonstrated good internal consistency and predictive validity.

One important disadvantage of the CD-Quest that limits its applicability in practice is its length: on average, it takes a participant 7 to 10 minutes to complete the full 15-item CD-Quest. This length may be prohibitive in a variety of contexts, such as when it is part of a battery that assesses a wide range of constructs, when it is administered repeatedly during an experiment with repeated measures, or in clinical settings.

1.1. Present study

We conducted a secondary analysis of previous CD-Quest datasets to develop and validate short-form models designed to reproduce the CD-Quest total score while: (1) using the responses from a smaller number of items (i.e., to reduce the amount of time required to administer the CD-Quest); and (2) assign small integer weights to each response (i.e., to ensure that the CD-Quest total score could still be computed by hand). Although these are common requirements for psychometric scales, there is no standard methodology to develop short-form scales with small integer weights. Existing scales are often developed by combining standard statistical methods with heuristics for rounding and variable selection (e.g., logistic regression with forward selection and unit weighting). We instead used a machine learning method called RiskSLIM (Risk-Calibrated Supersparse Linear Integer Model; Ustun & Rudin, 2019) to develop an *optimal* short-form of the CD-Quest. RiskSLIM uses modern optimization techniques to fit short-form models that perform well under operational constraints related to usability and interpretability – e.g., requiring models to use small integer weights for ease of hand scoring, assign positive or negative weights to variables, or to use a limited number of variables (e.g., Struck et al., 2017; Ustun et al., 2017; Zuromski et al., 2019).

Using RiskSLIM, we fit short-form models that could reproduce the total CD-quest score while using limited number of questions and assigning small integer weights to each response. We fit short-form models using data from a large sample of undergraduates in the US and evaluated their performance using data from five independent samples: two from undergraduates in Brazil, one from patients with depression in Brazil, one from patients with social anxiety disorder (SAD) in the US, and one from a mixed sample of psychiatric outpatients in Turkey.

2. Method

2.1. Participants and procedures

Archival data were used from six independent samples. A sample of 906 undergraduates was used as the training sample in which alternative short-form models were developed. We then tested model performance in reproducing total CD-Quest scores in the other samples. We focused on participants with no missing quantitative data (i.e., frequency and intensity ratings) on the CD-Quest (combined $N = 1,800$), given that 94.3% of the participants in the training sample and 82.7–100% in the test samples had complete data across all 15 items. From the full sample ($N = 1,895$), 49 (2.6%) were missing one item, 9 (0.5%) were missing two items, 10 (0.5%) were missing between 3–14 items, and 27 (1.4%) were missing all 15 items. The number of participants per sample excluded due to missing CD-Quest data is reported below. The proportion of participants excluded varied by sample, $\chi^2(df = 5, N = 1,895) = 46.92, p < .001$, with three samples missing more than expected (Samples 1 (undergraduates in the US), 3 (undergraduates in Brazil), and 4 (depressed patients in Brazil)), two missing fewer than expected (Samples 2 (undergraduates in Brazil) and 6 (psychiatric outpatients in Turkey)), and one missing as many as expected (Sample 5 (SAD patients in US)). However, missingness did not vary with age, Kruskal-Wallis $H(1) = 1.61, p = .21$, or sex, $\chi^2(df = 1, N = 1,895) = 0.41, p = .52$.

2.1.1. Sample 1: Undergraduates in the US

Initial fit of the short-form models was assessed in a sample of 906 undergraduates from Temple University, a large, urban, public university in the US previously reported in Morrison et al. (2015). We focused on participants with complete data, excluding 55 participants (5.7% of the full $N = 961$ sample) who were missing data on one or more items of the CD-Quest. Participants were students enrolled in a psychology course (64.8% female; $M_{age} = 20.7, SD = 3.7$, range = 18–60). Participants were compensated with course credit. Participation was completed online. First, participants provided informed consent. Then they completed a demographics questionnaire, followed by the CD-Quest. An additional 10 questionnaires were next administered but not analyzed in the current study. The study was approved by the IRB of Temple University.

2.1.2. Sample 2: Undergraduates in Brazil

The first of the validation samples comprised 182 Brazilian undergraduate medical and psychology students previously reported in de Oliveira et al. (2015a), again focusing on participants with complete data. Two participants (1.1% of the full $N = 184$ sample) were missing data on one item of the CD-Quest and were excluded from analysis. A majority of the participants were female (67.2%), and age was typical for a Brazilian college student sample ($M_{age} = 21.8, SD = 3.4$, range = 17–40). Participants provided informed consent, completed the CD-Quest and other questionnaires not analyzed here in their classrooms and then completed the questionnaires again 2–3 weeks later for retest analysis. Participants were not compensated. The study was approved by the IRB of the University Hospital at Federal University of Bahia.

2.1.3. Sample 3: Undergraduates in Brazil

The second validation sample consisted of an as-yet unpublished sample of 183 undergraduate students from the Federal University of Bahia, a public university, and from Ruy Barbosa College, a private university (Seixas et al., 2022). We again focused on students who provided complete data and we excluded 14 participants (7.1% of the full $N = 197$ sample) who were missing data on one or more items of the CD-Quest. Students were mostly enrolled in a course in human sciences or biological sciences, with a few enrolled in other courses and one student not reporting. The sample was predominantly female (75.7%), and mean age was in the mid-twenties ($M_{age} = 25.4, SD = 8.8$, range =

16-62). Participation occurred in classrooms, with participants first providing informed consent followed by completing a battery of questionnaires including the CD-Quest. Compensation was not provided to participants. The study was approved by the IRB of the Federal University of Bahia.

2.1.4. Sample 4: Depressed patients in Brazil

The third validation sample included 62 patients with depression participating and providing complete CD-Quest data in the randomized controlled trial comparing trial-based cognitive therapy, behavioral activation, and treatment-as-usual previously reported by Hemanny et al. (2019). We excluded 13 participants (17.3% of the full $N = 75$ sample) who were missing data on one or more items of the CD-Quest, with 10 of these participants missing all 15 items. Participants were majority female (61.0%) and older on average than the college student samples ($M_{age} = 39.5$, $SD = 11.0$, range = 16-60). Participants were recruited through mental health outpatient clinics, residency programs in psychiatry, and advertisements on the radio, newspapers, and the internet. Screening for the trial was conducted by a trained evaluator. Participants were eligible if they met the following criteria: current treatment with antidepressant medication for at least two months, age 18 to 60 years, diagnosis of major depressive disorder (MDD) according to DSM-IV (American Psychiatric Association [APA], 1994) or ICD-10 (World Health Organization, 2004) as assessed with the Mini International Neuropsychiatric Interview (MINI-plus; Amorim, 2000; Lecrubier et al., 1997). Further, eligible participants had to score higher than 15 on the Hamilton Depression Rating Scale (Hamilton, 1960) or higher than 20 on the Beck Depression Inventory (Beck et al., 1961; Cunha, 2001). Participants were ineligible if they endorsed any of the following: current use of mood stabilizing drugs; current participation in psychotherapy; high risk of suicide (according to the MINI-plus); diagnosis of bipolar disorder, psychotic disorders, or current substance abuse or dependence. Prior to screening, participants provided informed consent. Once deemed eligible, participants were randomized to one of two psychotherapy treatments (trial-based cognitive therapy or behavioral activation) or treatment-as-usual for the 12-week duration of the intervention. Participants randomized to treatment-as-usual could have their antidepressant monotherapy treatment modified at the discretion of their psychiatrists, whereas participants randomized to the psychotherapy treatments continued to use medication but were not to have their medication doses increased. Data were collected at baseline (pre-treatment), 6 weeks after starting treatment (mid-treatment), and 12 weeks after starting treatment (post-treatment). Only baseline/pre-treatment data were analyzed in the current study. In addition to completing the CD-Quest, participants completed the semi-structured diagnostic interview mentioned previously and other measures to evaluate treatment outcome not analyzed here. The study was approved by the IRB of the Health Sciences Institute at Federal University of Bahia. Participants were not compensated beyond receiving therapy as part of the randomized controlled trial.

2.1.5. Sample 5: Patients with SAD in the US

The fourth validation sample was comprised of 198 patients with a primary diagnosis of SAD who provided complete data on the CD-Quest. We excluded 11 participants (5.3% of the full $N = 209$ sample) who were missing data on one or more items of the CD-Quest. These patients came from two independent samples detailed below. A slight majority of participants were male (53.2%; $M_{age} = 30.2$, $SD = 8.8$, range = 18-64).

Seventy-eight of these patients were individuals seeking treatment for SAD at the Adult Anxiety Clinic of Temple University. Patients were eligible for the open trial of individual CBT for SAD if they had a primary diagnosis of generalized SAD according to DSM-IV or a primary diagnosis of SAD according to DSM-5 (APA, 2013). Exclusion criteria included the following: history of thought disorder, alcohol or substance dependence within the last six months, high risk of suicide, or current psychotherapy. Patients were permitted to have co-primary diagnoses,

so long as the comorbid difficulties were deemed well enough under control to focus on treatment of SAD. Concurrent psychotropic treatment was also permitted, and patients were encouraged to maintain their current dosage through the open trial. Patients were initially screened by telephone. Then they completed an in-person appointment, during which they provided informed consent, then were administered a diagnostic interview with the Anxiety Disorders Interview Schedule for the DSM-IV – Lifetime version (ADIS-IV-L; Di Nardo et al., 1994) or the Anxiety and Related Disorders Interview Schedule for the DSM-5 – Lifetime version (ADIS-5L; Brown & Barlow, 2014) by a trained doctoral-level graduate student. Patients then completed pre-treatment assessments including the CD-Quest and other measures not analyzed here before they began an open trial of individual CBT. Therapy was provided on a sliding scale and patients were not provided any compensation for participating. The study was approved by the Temple University IRB.

One hundred twenty of these patients were recruited to participate in a 3-arm randomized controlled trial comparing cognitive behavioral group therapy (CBGT) to mindfulness-based stress reduction (MBSR) and a waitlist control, originally reported in Goldin et al. (2016). In addition, psychometric properties of the full CD-Quest from these participants were reported in Kaplan et al. (2017). Patients were eligible for the trial if they (1) met DSM-IV criteria for a principal diagnosis of generalized SAD according to diagnostic interview with the ADIS-IV-L, (2) endorsed greater than moderate fear in 5 or more different social situations on the ADIS-IV-L, and (3) scored over 60 on the Liebowitz Social Anxiety Scale – Self-Report (LSAS-SR; Fresco et al., 2001; Liebowitz, 1987). Exclusion criteria included the following: treatment with pharmacotherapy or psychotherapy in the past year; significant experience with mindfulness (i.e., enrollment in any previous MBSR course, participation in any long-term meditation retreats, or history of regular meditation practice); participation in CBT for any anxiety disorder in the last 2 years; history of neurological disorders, cardiovascular disorders, thought disorders, eating disorders, or bipolar disorders; current substance or alcohol abuse/dependence; major depressive episode in the previous month; current posttraumatic stress disorder or having weekly or daily flashbacks that interfered with functioning; or current co-primary obsessive-compulsive disorder or more than one hour of obsessive or compulsive symptoms each day that interfere with daily life. Patients were recruited through community listings and referrals from clinicians. They were initially screened online, followed by telephone interview, and then an in-person diagnostic assessment with the ADIS-IV-L. The CD-Quest and other self-report assessments were completed online thereafter, with the option to complete the questionnaires in more than one sitting to reduce fatigue. Patients also completed a functional magnetic resonance imaging assessment and other behavioral tasks prior to randomization ($n = 108$) to 12 weeks of CBGT, MBSR, or waitlist. Therapy was provided free of charge and patients were not provided any additional compensation for participating in any baseline, mid-treatment, or post-treatment assessments. Follow-ups were conducted every three months post-treatment via online self-report questionnaires. Modest financial compensation was provided for the 12-month follow-up assessment which additionally included a behavioral assessment. The study was approved by the Stanford University IRB.

2.1.6. Sample 6: Psychiatric outpatients in Turkey

The fifth validation sample was comprised of 269 psychiatric outpatients presenting to three tertiary healthcare services in Ankara and Tokat, Turkey. There were no missing data on the CD-Quest in this sample. Psychometric properties of the full CD-Quest from these participants were previously reported in Batmaz et al. (2015). Participants were predominantly female (61.0%), with an average age in the mid-thirties ($M_{age} = 36.4$, $SD = 12.5$, range = 18-65). Participants predominantly presented for treatment with depressive and anxiety symptoms and were permitted to be taking psychotropic medications.

Exclusion criteria included the following: diagnosis with psychotic disorders, bipolar disorders, organic mental disorders, dementia, or intellectual disability; having an uncontrolled medical or neurological disorder unless mild in nature; high risk of suicide at time of intake interview; history of head trauma, recent brain surgery, or electroconvulsive therapy; or current participation in CBT. After providing informed consent and participating in a diagnostic interview with the MINI (Sheehan et al., 1998), participants completed the CD-Quest by paper-and-pencil, along with other self-report questionnaires not analyzed here. Participants were not provided compensation. The study was approved by the ethics committee of Tokat Gaziosmanpaşa University.

2.2. Cognitive Distortions Questionnaire

The Cognitive Distortions Questionnaire (CD-Quest; de Oliveira, 2014, 2015a) consists of pairs of questions about frequency and intensity of 15 common cognitive distortions over the past week. The item development procedure has been reported previously (see de Oliveira et al., 2015, and Morrison et al., 2015). The questionnaire begins by presenting instructions that include a brief definition of thoughts and thinking errors (i.e., cognitive distortions) followed by a vignette depicting one type of thinking error. The instructions explain how the person in the vignette should respond to the CD-Quest item on that cognitive distortion. After these instructions, the questionnaire presents the names and brief descriptions of 15 common cognitive distortions. Each description is followed by two illustrative thoughts/beliefs characterized by the cognitive distortion. For example, the second item is entitled “Fortune telling (also called catastrophizing)” and is described with the statement, “I predict the future in negative terms and believe that what will happen will be so awful that I will not be able to stand it.” The two illustrative thoughts/beliefs are, “I will fail and this will be unbearable” and “I’ll be so upset that I won’t be able to concentrate for the exam.” The respondent is then asked to enter their own example of a thought/belief characterized by this cognitive distortion and then to rate the frequency and intensity of their belief in this distortion over the past week. Response options for each distortion are presented in a table with 4 columns and 3 rows. The columns contain 4 response options that include one indicating that the cognitive distortion did not occur during the previous week and the other 3 indicating frequency of occurrence: occasional (1-2 days during the past week), much of the time (3-5 days during the past week), and almost all of the time (6-7 days during the past week). The rows contain response options for intensity, with the prompt “I believed it...” and the response options: a little (up to 30%), much (31% to 70%), or very much (more than 70%). For distortions that do not occur in the previous week, there is no intensity rating. The cross-classification of frequency-intensity for each of the 15 cognitive distortions is scored on a 0-5 response scale where 0 indicates that the distortion did not occur and scores of 1-5 are created by subtracting 1 from the sum of scores across rows and columns, where frequency and intensity are both coded in the range 1-3. The CD-Quest produces total scores for frequency (0-45), intensity (0-45), and the composite (0-75). We did not use write-in responses in the current analyses. The full CD-Quest is presented in Appendix A.

In Samples 1 and 5, the English version of the CD-Quest was used. In Samples 2, 3, and 4, the Brazilian Portuguese version was used. In Sample 6, a Turkish version was used. The English and Brazilian Portuguese versions were constructed at the same time by the fifth author (I. R.O.) (de Oliveira, 2014, 2015a, 2015b). The Turkish version, developed by Batmaz et al. (2015) was first translated from English by that study’s first author (S.B.), then the translated scale was back-translated by two bilingual experts in the field, and all translations were compared with the original English version of the scale. After reviewing the original and translated versions, a final consensus version was adopted.

2.3. Model development

2.3.1. Data processing

We were interested in determining how well CD-Quest total composite scores could be reproduced by reducing the number of items (i.e., cognitive distortions) included in the scale. We fit short-form models using the data from Sample 1 and then used the remaining samples to determine the validity of these short-form models. We chose Sample 1 as the training sample because it contained the largest number of respondents. Our decision to train the model using data collected in a single study as opposed to using a training sample selected from all studies was meant to avoid the risk of combining data from potentially heterogeneous sources and introducing dependence between the training sample and the test samples. A significant benefit of this decision was that the short-forms developed in the training sample could be validated in multiple distinct data sets to examine generalizability of the short-forms.

To ensure that a short-form model would extract as much information as possible from responses to the short-form questions, we converted the 0-5 composite response score for each of the 15 cognitive distortions (CDs) into a set of 5 nested 0-1 binary threshold variables of the form: Variable CDj1 was coded 1 if the 0-5 composite score for CDj ≥ 1 and was otherwise coded 0; Variable CDj2 was coded 1 if the 0-5 composite score for CDj ≥ 2 and was otherwise coded 0; Variable CDj3 was coded 1 if the 0-5 composite score for CDj ≥ 3 and was otherwise coded 0; Variable CDj4 was coded 1 if the 0-5 composite score for CDj ≥ 4 and was otherwise coded 0; and Variable CDj5 was coded 1 if the 0-5 composite score for CDj = 5 and was otherwise coded 0. Thus, our final dataset included the original CD-Quest total composite score along with 75 input variables: 5 nested 0-1 binary threshold variables for each of 15 items.

2.3.2. Model estimation

RiskSLIM¹ is a procedure for estimating constrained optimal prediction models according to some objective outcome criterion with constraints imposed on the number of predictors included in the model and allowing patterned constraints to be imposed both on these predictors (in our case, allowing either all 5 or none of the 5 variables representing a single item to be included in the model) and on their slopes (in our cases, requiring slopes to be whole numbers in the range 0-5). We fit each short-form model to optimize the mean-squared error between individual-level predicted score and the original total composite score. This represents a minor departure from prior uses of RiskSLIM, in which optimal short-form models were used to predict probabilities on yes-no dichotomous outcomes by optimizing likelihoods under the logistic distribution rather than minimizing mean-squared error, but in both cases the same techniques are used to optimize convex loss functions (Ustun & Rudin, 2019). For the sake of clarity, we report model performance in terms of R^2 as a standardized value of mean-squared error, where $R^2 = 0\%$ means that the RiskSLIM model is unrelated to the observed original CD-Quest score and $R^2 = 100.0\%$ means that the RiskSLIM score can perfectly reproduce CD-Quest scores. We fit short-form models for between 4 to 9 items, deciding *a priori* to stop at 9 items based on our aim of developing a significantly shorter version.

3. Results

3.1. Overall model fit

The R^2 for the short-form models in the training sample ranged between 80.2% for the 4-item model and 91.7% for the 9-item model. The

¹ Python code for RiskSLIM is publicly available at <https://github.com/ustunb/risk-slim>

models exhibited excellent generalization across the test samples, with R^2 close to those in the training sample (Table 1). The exception is the 4-item model in Sample 4 (patients with depression in Brazil), for which R^2 is notably lower than in the other samples. However, this discrepancy disappears in the 5-item model and does not reappear in the 6- to 9-item models. The strong generalizability of results across the test samples is striking given the differences across samples in demographics, nationality, language, and severity of psychopathology.

Prediction accuracy of composite total CD-Quest scores can also be seen by graphing the association between observed (X axis) and predicted (Y axis) composite CD-Quest scores for each of the models within samples (Supplemental Figures 1-6). The most striking change is seen in Sample 4, for which $R^2 = 63.7\%$ for the 4-item model, 78.2% for the 5-item model, and up to 90.9% for the 9-item model (Table 1). The 4-item model had relatively low prediction accuracy for CD-Quest total scores between 0-20 and underestimated higher total scores. With the 5-item model and higher, prediction accuracy in the lower range improved and underestimation in the higher range reduced. Inspection of comparable plots for the other models and samples shows that all RiskSLIM models, even though based on small numbers of CDs, have generally very good estimation.

Two aspects of these results are especially noteworthy. First, part-whole reliability is typically assessed, similar to split-half reliability, with the correlation between the part and the whole rather than by R^2 . This means that the reliability of the short-form scales relative to the total scale in the test samples averaged between .90 for the 5-item ultra-short form and .96 for the 9-item short-form. Second, we would expect these part-whole measures to become very large as the number of items in the short-form scales approaches the number in the full scale, as the content is increasingly overlapping. This makes the .90 average reliability of the 5-item ultra-short form relative to the full scale especially impressive given that these scales have only one-third of their items in common.

3.2. RiskSLIM item selection and scoring

Table 2 presents the cognitive distortions selected in each RiskSLIM model. CD8 (Mind reading) and CD11 (Should statements) were selected by RiskSLIM in all models. CD10 (Personalization) was selected in all but the 9-item model, CD9 (Overgeneralization) in the 3 models with the fewest CDs, CD6 (Magnification/minimization) in the models with intermediate numbers of CDs, and CDs 14 and 15 (What if? and Unfair comparisons) in the models with the highest numbers of CDs. The only CD not selected by RiskSLIM in any model was CD12 (Jumping to conclusions).

Indices of overall model fit and inspection of plots suggest that short-form models reproduce the full CD-Quest scores with good (5-item ultra-short-form) to excellent (9-item short-form) accuracy. As a short-form of

Table 1
 R^2 of RiskSLIM Models Reproducing CD-Quest Total Composite Scores in the Training Sample and Test Samples.

	Number of items/CDs in the Model							N
	4	5	6	7	8	9		
Sample 1 (training)	80.2	83.7	85.7	89.4	90.7	91.7	906	
Sample 2	76.1	82.5	81.2	82.0	84.8	90.4	182	
Sample 3	76.0	79.5	87.4	89.7	89.5	90.8	183	
Sample 4	63.7	78.2	85.5	87.4	90.0	90.9	62	
Sample 5	78.1	82.3	86.6	86.1	89.3	93.0	198	
Sample 6	83.6	85.5	90.3	90.8	91.3	93.6	269	

Note. CD = cognitive distortion; Sample 1 = undergraduates in US; Samples 2 and 3 = undergraduates in Brazil; Sample 4 = patients with depression in Brazil; Sample 5 = patients with social anxiety disorder in US; Sample 6 = psychiatric outpatients in Turkey. Standard deviations of R^2 for true score values in the range here are between approximately 5.0 for a sample of 62 and less than 2.0 for a sample of 906 (Cramer, 1987).

Table 2
The CD-Quest Cognitive Distortions included in the RiskSLIM models.

CD	Number of items/CDs in the Model					
	4	5	6	7	8	9
1 Dichotomous thinking				X	X	
2 Fortune telling			X			
3 Discounting the positive						X
4 Emotional reasoning						X
5 Labeling						X
6 Magnification/minimization		X	X	X	X	
7 Selective abstraction						X
8 Mind reading	X	X	X	X	X	X
9 Overgeneralization	X	X	X			X
10 Personalization	X	X	X	X	X	
11 Should statements	X	X	X	X	X	X
12 Jumping to conclusions						
13 Blaming			X		X	
14 What if?				X	X	X
15 Unfair comparisons				X	X	X

Note. CD = cognitive distortion.

the CD-Quest, we recommend the 9-item model (CD-Quest-S9), which has an R^2 of 91.7% in the training sample and $90.4\% - 93.6\%$ in the test samples. As an ultra-short form of the CD-Quest, we recommend the 5-item model (CD-Quest-S5), which has an R^2 of 83.7% in the training sample and $78.2\% - 85.5\%$ in the test samples. In Appendices B and C, we present these short forms with the RiskSLIM weights for reproducing the estimated CD-Quest total score embedded within the response table for each item, which matches the formatting of the original CD-Quest and may ease the burden of hand-scoring.

The full set of RiskSLIM weights used for scoring the various short-form models are provided in Supplemental Table 1. Weights are provided for each nested variable dichotomy. For example, for scoring the item "Discounting the positive" in the 9-item model, a composite response of 4 would be scored as 6 (i.e., looking at Supplemental Table 1 shows us the points for each threshold dichotomy as follows: 4 is greater than 1 (2 points), greater than 2 (2 points), greater than 3 (2 points), equal to 4 (0 points), not greater than 5 (no points)). Similar scoring based on the weights provided in the table would be done for the remaining 8 items in the 9-item model and then these 9 scores should be summed together with the offset value for the model (i.e., 1 for the 9-item model) to provide the total estimated composite score of the full CD-Quest.

4. Discussion

Within cognitive behavioral approaches, cognitive distortions are viewed as playing a central role in the development, maintenance, and treatment of many mental disorders. Therefore, reliable and valid assessment of cognitive distortions is of great importance. Nevertheless, few measures have been designed for this purpose, and only one such measure, the CD-Quest (de Oliveira et al., 2015), has undergone significant psychometric evaluation. Given evidence of the strong psychometric properties of the CD-Quest, we sought to develop short forms of the CD-Quest to further enable its use by researchers and clinicians who may be limited in time and resources. This is important because the full scale takes 7-10 minutes to complete. The 5-item ultra-short form developed here reduces the number of items in the original 15-item scale by two-thirds and cuts administration time to 3.5-5 minutes while retaining .90 reliability relative to the full scale. The 9-item short-form developed here, in comparison, reduces the number of items by 40% and cuts administration time to 5.5-7 minutes while retaining .96 reliability relative to the full scale.

We developed and externally validated various short-form models of the CD-Quest using RiskSLIM (Ustun & Rudin, 2019), a machine-learning approach that optimized the prediction of the full CD-Quest total score with respect to R^2 based on fewer items and with

small integer weights. The short forms were first fit in a US undergraduate sample ($N = 906$) and then validated in five separate and demographically distinct samples, including two samples of undergraduate students in Brazil ($Ns = 182$ & 183), a sample of patients with depression in Brazil ($N = 62$), a sample of patients with SAD in the US ($N = 198$), and a sample of psychiatric outpatients in Turkey ($N = 269$). This kind of external validation in diverse independent samples is recognized as superior to either cross-validation in a training sample or comparison of training and test subsamples in a single large development sample (König et al., 2007; Youngstrom et al., 2018).

Model fit for the various short-form models, ranging from 4 to 9 of the original 15 CD-Quest items, was good-to-excellent both in the training sample and in all five validation samples, with one exception: the 4-item ultra-short form model did not generalize well to the sample of patients with depression in Brazil ($R^2 = 63.7\%$). However, fit improved significantly even in that sample for the 5-item model ($R^2 = 78.2\%$), with fit in this and the other validation samples ($R^2 = 78.2\% - 85.5\%$) similar to that in the training sample ($R^2 = 83.7\%$). Model fit continued to improve in all samples in models with more items, up to the 9-item model which showed excellent fit in all samples ($R^2 = 90.4\% - 93.6\%$). Therefore, we recommend the 9-item model as a short-form of the CD-Quest (CD-Quest-S9) and the 5-item model as an ultra-short form of the CD-Quest (CD-Quest-S5). We estimate the short forms to save 3 to 5 minutes in administration time, which is significant for an original scale that takes approximately 10 minutes to complete.

One unique benefit of using RiskSLIM to develop short forms of the CD-Quest is that it provides integer weights assigned to each response of the short form, which enables the user to easily compute a predicted CD-Quest total score by hand. This calculated total score can then be compared to total scores on the full CD-Quest to track therapy progress or to compare across research samples. In Appendices B and C, we provided the CD-Quest-S9 and CD-Quest-S5 with integer weights integrated into the questionnaires to ease the process of hand-scoring and in Supplemental Table 1, we provided the integer weights for all tested short-form models. Hand scoring has potential value for purposes of facilitating quick clinical decisions if the CD-Quest is used for precision treatment planning. However, the use of integer weights constrains the ability to optimize fit, which means that when scoring can be computerized it might be better to use computer adaptive testing (CAT) methods based on item response theory to develop an optimal short-form scale.

The current study had several strengths, including the large and diverse samples and the consistency of model fit across these samples. Importantly, the samples differed in clinical status (as well as clinical diagnosis across the clinical samples), geographic region, and language, thus providing strong evidence of generalizability. In addition, there were several limitations. First, the use of R^2 as the objective function to be optimized in developing the short-form and ultra-short form might be suboptimal, especially if a clinical decision is needed that requires determination of whether a patient is above or below a given threshold on the full CD-Quest. If the latter is the case, a more useful approach would be to use item response theory or a modification of RiskSLIM that attempts to minimize classification error, perhaps with a weight for a higher importance of false negatives than false positives, using a logistic link function to predict the dichotomous outcome based on the total score and optimizing an objective function such as AUC.

Second, psychometric properties of the recommended short forms were not evaluated. However, it is important to recognize that the optimization scheme trades off internal consistency for predictive validity by selecting the minimally redundant items, making it uninformative to use a conventional assessment of internal consistency when evaluating psychometric properties.

Third, although there were both nonclinical and clinical samples in the current study, it could be argued that all participants were relatively high in education and/or socioeconomic status, and may also have had relatively high levels of conscientiousness. Specifically, nonclinical

samples were comprised of only college students and clinical samples were comprised of only treatment-seeking individuals with primary anxiety and/or depression. Research is needed to evaluate whether the strong psychometric properties of the CD-Quest generalize to individuals lower in education and/or socioeconomic status, or to individuals with primary attention problems, impulsivity, or oppositional attitudes or behaviors. This is especially important given that the CD-Quest is a cognitively complex scale.

Fourth, as is typical of research conducted in clinical samples, there were relatively stringent eligibility criteria for recruitment of participants in the parent studies of our three clinical samples. Some of these criteria differed by sample (e.g., inclusion requirement of a primary diagnosis of MDD in the Brazil sample versus SAD in the US sample), but other criteria were uniform across the three samples and might exclude a meaningful subset of patients who seek CBT. Clinicians planning to use the CD-Quest or its short-forms should take this lack of evidence into consideration when drawing inferences about patients excluded from previous research, such as those with psychotic disorders, bipolar disorders, or acute suicidality.

Finally, a small number ($n = 95$; 5.0%) of participants were excluded due to missing data on one or more items of the CD-Quest. Although age and sex were not associated with missingness, sample was, which suggests various possible explanations for missingness, such as differences in study administration or clinical severity. However, the pattern of missing data across our nonclinical versus clinical samples does not bear out the latter idea, since one clinical sample had more-than-expected missing data (Sample 4), another had as expected (Sample 5), and the third had the least missing data across all samples (Sample 6). Although results may have changed somewhat if we had used a different method to handle missing data, such as multiple imputation, we chose not to do so based on the low proportion of cases with any item-missing data and the fact that the associations of observed data with item missingness were modest.

In sum, we used a machine-learning algorithm to test and validate various shortened versions of the 15-item CD-Quest. Across six large, diverse nonclinical and clinical samples, we found evidence of good-to-excellent model fit using 5- to 9-item short-forms of the CD-Quest with optimal integer scoring rules to reproduce the total score of the original CD-Quest. We therefore recommend a 9-item short version and 5-item ultra-short version of the CD-Quest to users who need an efficient tool to assess cognitive distortions.

Funding

This research was supported by NIMH Grant R01 MH092416, awarded to James J. Gross. ClinicalTrials.gov identifier: NCT02036658. Ekaterina Sadikova was supported by NIMH Grant 2 T32 MH 17119-34.

Declaration of Competing Interest

In the past 3 years, Dr. Kessler was a consultant for Datastat, Inc., RallyPoint Networks, Inc., Sage Pharmaceuticals, and Takeda. Dr. Heimberg was a consultant for VistaGen Therapeutics, Inc. Dr. de Oliveira created the CD-Quest. None of the remaining authors have any conflicts of interest to report.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.janxdis.2022.102624](https://doi.org/10.1016/j.janxdis.2022.102624).

References

- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders (4th edition)*. Washington, DC: Author.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. (<https://doi.org/10.1176/appi.books.9780890425596>).

- Amorim, P. (2000). Mini internacional neuropsychiatric interview (MINI): Validação de entrevista breve para diagnóstico de transtornos mentais. *Brazilian Journal of Psychiatry*, 22(3), 106–115. <https://doi.org/10.1590/S1516-4446200000300003>
- Batmaz, S., Kochbiyik, S., & Yuncu, O. A. (2015). Turkish version of the Cognitive Distortions Questionnaire: Psychometric properties. *Depression Research and Treatment*, 2015, 1–8. <https://doi.org/10.1155/2015/694853>
- Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. Oxford, England: International Universities Press.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4(6), 561–571.
- Brown, T. A., & Barlow, T. H. (2014). *Anxiety and related disorders interview schedule for DSM-5, adult and lifetime version: Clinician manual*. New York, NY: Oxford University Press.
- Butler, R. M., O'Day, E. B., Swee, M. B., Horenstein, A., & Heimberg, R. G. (2021). Cognitive behavioral therapy for social anxiety disorder: Predictors of treatment outcome in a quasi-naturalistic setting. *Behavior Therapy*, 52(2), 465–477. <https://doi.org/10.1016/j.beth.2020.06.002>
- Covin, R., Dozois, D. J. A., Ogniewicz, A., & Seeds, P. M. (2011). Measuring cognitive errors: Initial development of the cognitive distortions scale (CDS). *International Journal of Cognitive Therapy*, 4(3), 297–322. <https://doi.org/10.1521/ijct.2011.4.3.297>
- Cramer, J. S. (1987). Mean and variance of R^2 in small and moderate samples. *Journal of Econometrics*, 35, 253–266. [https://doi.org/10.1016/0304-4076\(87\)90027-3](https://doi.org/10.1016/0304-4076(87)90027-3)
- Cunha, J. A. (2001). *Manual da Versão em Português das Escalas Beck*. Casa do Psicólogo: São Paulo (SP).
- de Oliveira, I. R. (2014). Trial-based cognitive therapy (TBCT): A new cognitive-behavior therapy approach. In I. R. de Oliveira, T. Schwartz, & S. N. Stahl (Eds.), *Integrating psychotherapy and psychopharmacology: A handbook for clinicians* (pp. 24–65). New York: Routledge.
- de Oliveira, I. R. (2015a). Introducing the Cognitive Distortions Questionnaire. In I. R. de Oliveira (Ed.), *Trial-based cognitive therapy: A manual for clinicians* (pp. 25–40). New York: Routledge.
- de Oliveira, I. R. (2015b). Terapia cognitiva processual (TCP): Uma nova abordagem de terapia cognitivo-comportamental. In I. R. de Oliveira, T. Schwartz, & S. Stahl (Eds.), *Integrando Psicoterapia e Psicofarmacologia* (pp. 26–65). Porto Alegre (RS): Artmed.
- de Oliveira, I. R., Seixas, C., Osório, F. L., Crippa, J. A. S., de Abreu, J. N., Menezes, I. G., Pidgeon, A., Sudak, D., & Wenzel, A. (2015). Evaluation of the psychometric properties of the Cognitive Distortions Questionnaire (CD-Quest) in a sample of undergraduate students. *Innovations in Clinical Neuroscience*, 12(7-8), 20–27.
- Di Nardo, P. A., Brown, T. A., & Barlow, D. H. (1994). *Anxiety Disorders Interview Schedule for DSM-IV: Lifetime version (ADIS-IV-L)*. New York, NY: Oxford University Press.
- Fresco, D. M., Coles, M. E., Heimberg, R. G., Liebowitz, M. R., Hami, S., Stein, M. B., & Goetz, D. (2001). The Liebowitz social anxiety scale: A comparison of the psychometric properties of self-report and clinician-administered formats. *Psychological Medicine*, 31(6), 1025–1035. <https://doi.org/10.1017/S0033291701004056>
- Goldin, P. R., Morrison, A., Jazaieri, H., Brozovich, F., Heimberg, R., & Gross, J. J. (2016). Group CBT versus MBSR for social anxiety disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 84(5), 427–437. <https://doi.org/10.1037/ccp0000092>
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, 23(1), 56–62. <https://doi.org/10.1136/jnnp.23.1.56>
- Hemanny, C., Carvalho, C., Maia, N., Reis, D., Botelho, A. C., Bonavides, D., Seixas, C., & de Oliveira, I. R. (2019). Efficacy of trial-based cognitive therapy, behavioral activation and treatment as usual in the treatment of major depressive disorder: Preliminary findings from a randomized clinical trial. *CNS Spectrums*, 25(4), 535–544. <https://doi.org/10.1017/S1092852919001457>
- Kaplan, S. C., Morrison, A. S., Goldin, P. R., Olino, T. M., Heimberg, R. G., & Gross, J. J. (2017). The cognitive distortions questionnaire (CD-Quest): Validation in a sample of adults with social anxiety disorder. <https://doi.org/10.1007/s10608-017-9838-9>
- König, I. R., Malley, J. D., Weimar, C., Diener, H. C., Ziegler, A., & German Stroke Study, C. (2007). Practical experiences on the necessity of external validation. *Statistics in Medicine*, 26, 5499–5511. <https://doi.org/10.1002/sim.3069>
- Kostoglou, S. L., & Pidgeon, A. M. (2016). The cognitive distortions questionnaire: Psychometric validation for an Australian population. *Australian Journal of Psychology*, 68(2), 123–129. <https://doi.org/10.1111/ajpy.12101>
- Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., Janavs, J., & Dunbar, G. C. (1997). The mini international neuropsychiatric interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry*, 12(5), 224–231. [https://doi.org/10.1016/S0924-9338\(97\)83296-8](https://doi.org/10.1016/S0924-9338(97)83296-8)
- Lefebvre, M. F. (1981). Cognitive distortion and cognitive errors in depressed psychiatric and low back pain patients. *Journal of Consulting and Clinical Psychology*, 49(4), 517–525. <https://doi.org/10.1037/0022006x.49.4.517>
- Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*, 22, 141–173.
- Morrison, A. S., Potter, C. M., Carper, M. M., Kinner, D. G., Jensen, D., Bruce, L., Wong, J., de Oliveira, I. R., Sudak, D. M., & Heimberg, R. G. (2015). The Cognitive Distortions Questionnaire (CD-Quest): Psychometric properties and exploratory factor analysis. *International Journal of Cognitive Therapy*, 8(4), 287–305. <https://doi.org/10.1521/ijct.2015.8.4.287>
- Qian, L., Liu, L., Chen, M., Wang, S., Cao, Z., & Zhang, N. (2020). Reliability and validity of the Chinese version of the Cognitive Distortions Questionnaire (CD-Quest) in college students. *Medical Science Monitor*, 26, e926786–1–e926786-8. <https://doi.org/10.12659/MSM.926786>
- Seixas, C., Vieira, R., Pires, P.P., Butler, R.M., & de Oliveira, I.R. (2022). Cognitive Distortions Questionnaire (CD-Quest): Validation and psychometric properties in a sample of Brazilian undergraduate students. Manuscript submitted for publication.
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The mini-international neuropsychiatric interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, 59(20), 22–33.
- Struck, A. F., Ustun, B., Ruiz, A. R., Lee, J. W., LaRoche, S. M., Hirsch, L. J., Gilmore, E. J., Vlachy, J., Haider, H. A., Rudin, C., & Westover, M. R. (2017). Association of an electroencephalography-based risk score with seizure probability in hospitalized patients. *JAMA Neurology*, 74(12), 1419–1424. <https://doi.org/10.1001/jamaneurol.2017.2459>
- Ustun, B., & Rudin, C. (2019). Learning optimized risk scores. *Journal of Machine Learning Research*, 20, 1–10. (<http://arxiv.org/abs/1610.00168>).
- Ustun, B., Adler, L. A., Rudin, C., Faraone, S. V., Spencer, T. J., Berglund, P., Gruber, M. J., & Kessler, R. C. (2017). The World Health Organization Adult Attention-Deficit/Hyperactivity Disorder Self-Report Screening Scale for DSM-5. *JAMA Psychiatry*, 74(5), 520–526. <https://doi.org/10.1001/jamapsychiatry.2017.0298>
- World Health Organization (2004). *ICD-10: International statistical classification of diseases and related problems*. (10th revision, 2nd ed.). (<https://apps.who.int/iris/handle/10665/42980>).
- Youngstrom, E. A., Halverson, T. F., Youngstrom, J. K., Lindhiem, O., & Findling, R. L. (2018). Evidence-based assessment from simple clinical judgments to statistical learning: Evaluating a range of options using pediatric bipolar disorder as a diagnostic challenge. *Clinical Psychological Science*, 6, 234–265. <https://doi.org/10.1177/2167702617741845>
- Yurica, C.L. (2002). *Inventory of Cognitive Distortions: Development and validation of a psychometric test for the measurement of cognitive distortions*. Unpublished doctoral dissertation, Philadelphia College of Osteopathic Medicine, Pennsylvania.
- Zuromski, K. L., Ustun, B., Hwang, I., Kean, T. M., Marx, B., Stein, M. B., Ursano, R. J., & Kessler, R. C. (2019). Developing an optimal short-form of the PTSD Checklist for DSM-5 (PCL-5). *Depression & Anxiety*, 36(9), 790–800. <https://doi.org/10.1002/da.22942>