



## Faculty Publications

---

2020-04-08

# Establishment of a Reliable Change Index for the GAD-7

Thomas Bischoff  
*Pfeiffer University*

Shayne R. Anderson  
*Brigham Young University - Provo, shayne\_anderson@byu.edu*

Joy Heafner

Rachel Tambling  
*University of Connecticut - Storrs*

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>



Part of the [Other Social and Behavioral Sciences Commons](#)

## Original Publication Citation

Bischoff, Thomas et al. Establishment of a Reliable Change Index for the GAD-7. *Psychology, Community & Health*, [S.l.], v. 8, n. 1, p. 176-187, apr. 2020.

---

## BYU ScholarsArchive Citation

Bischoff, Thomas; Anderson, Shayne R.; Heafner, Joy; and Tambling, Rachel, "Establishment of a Reliable Change Index for the GAD-7" (2020). *Faculty Publications*. 4117.  
<https://scholarsarchive.byu.edu/facpub/4117>

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).



## Empirical Articles

# Establishment of a Reliable Change Index for the GAD-7

Estabelecimento de um Índice de Mudança Confiável para o GAD-7

Thomas Bischoff<sup>a</sup>, Shayne R. Anderson<sup>b</sup>, Joy Heafner<sup>c</sup>, Rachel Tambling<sup>d</sup>

[a] Marriage and Family Therapy, Pfeiffer University, Charlotte, NC, USA. [b] Marriage and Family Therapy, Brigham Young University, Provo, UT, USA. [c] Private Practice. [d] Human Development and Family Studies, University of Connecticut, Storrs, CT, USA.

## Abstract

**Aim:** It is increasingly important for mental healthcare providers and researchers to reliably assess client change, particularly with common presenting problems such as anxiety. The current study addresses this need by establishing a Reliable Change Index of 6 points for the GAD-7.

**Method:** Sample size included 116 online community participants using Amazon's Mechanical Turk (MTurk) and archival data for 332 clinical participants. Participants completed measures of the GAD-7 and the MDI in 2 rounds. Using previously established cutoff scores and Jacobson and Truax's (1991) method, we establish a Reliable Change Index which, when applied to 2 administrations of the GAD-7, indicates if a client has experienced meaningful change.

**Results:** For the GAD-7, the mean score for the clinical sample was 10.57. For the community sample at Time 1, the mean score was 4.14. A Pearson's correlation was computed to assess the 14-28-day test-retest reliability of the GAD-7,  $r(110) = .87$ , indicating good test-retest reliability.

**Conclusion:** Using the RCI equation, this resulted in an RCI of 5.59. For practical use the RCI would be rounded to 6.

**Keywords:** reliable change index, clinical significance, clinically meaningful change, anxiety, GAD-7

## Resumo

**Objetivo:** É de extrema importância que os profissionais de saúde mental e investigadores consigam avaliar de forma fidedigna a mudança do cliente, especialmente no que diz respeito a problemas comuns como é o caso da ansiedade. O presente estudo aborda esta necessidade estabelecendo um Índice de Mudança Confiável de 6 pontos para o GAD-7.

**Método:** A amostra incluiu 116 participantes de uma comunidade online utilizando a Amazon's Mechanical Turk (MTurk) juntamente com dados de arquivo de 332 participantes clínicos. Os participantes completaram os instrumentos de avaliação GAD-7 e MDI em 2 momentos. Utilizando scores de cutoff previamente estabelecidos e o método de Jacobson e Truax's (1991), foi estabelecido o Índice de Mudança Confiável (RCI) que, quando aplicado a dois administradores do GAD-7 indica se um cliente experienciou uma mudança significativa.

**Resultados:** Para o GAD-7, o mean score para a amostra clínica foi de 10.57. Relativamente à amostra comunitária no primeiro momento, o mean score foi de 4.14. Foi utilizada a correlação de Pearson para avaliar a fiabilidade teste-reteste de 14-28 dias do GAD-7,  $r(110) = .87$ , indicando uma fiabilidade de teste-reteste boa.

**Conclusão:** Utilizando a equação RCI, o resultado diz respeito a um RCI de 5.59. Para que exista uma utilização prática, é necessário que o RCI seja próximo de 6.

**Palavras-Chave:** índice de mudança confiável, significância clínica, mudança clinicamente significativa ansiedade, GAD-7

Psychology, Community & Health, 2020, Vol. 8(1), 176–187, <https://doi.org/10.5964/pch.v8i1.309>

Received: 2018-11-06. Accepted: 2019-07-31. Published (VoR): 2020-04-08.

Handling Editor: Marta Matos, ISCTE - Instituto Universitário de Lisboa CIS - IUL, Lisbon, Portugal

\*Corresponding author at: Division of Applied Health Sciences, 1515 Mockingbird Lane, Charlotte, NC 28209, USA. Phone: 801-389-6404, E-mail: [tommybischoff@gmail.com](mailto:tommybischoff@gmail.com)



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is important to both clinicians and researchers to measure the effectiveness of mental health treatment. The community in which they serve benefits from solid research and effective implementation of findings. Furthermore, third party payers, such as insurance companies, are also invested in quantifiably measuring client change as a result of mental health interventions. Focus on appropriate mental health measurements is increasing as knowledge and services are expanding all over the world (e.g. instruments being translated and validated in multiple languages; [Carvalho, Marques, Ferreira, & Lima, 2016](#); [Dias, Silva, Maroco, & Campos, 2015](#)).

One method of measuring the effects of treatment is to administer questionnaires to clients to first assess their baseline state on some construct, and then to re-administer the questionnaire at a later time point to ascertain if there has been any change. Many instruments have been validated through assessment of its psychometric properties ([Losoi et al., 2013](#); [Marques et al., 2013](#); [Pimenta, Leal, & Maroco, 2012](#)) and comparison with other established measurements ([Barry, Folkard, & Ayliffe, 2014](#)). One common construct of interest to clients, clinicians, and third parties is anxiety, which is one of the most frequently presented problems that clients report when they seek therapy ([Heafner, Silva, Tambling, & Anderson, 2016](#)). Not only is anxiety a common presenting problem, it is a risk factor for a variety of physical ailments including cardiovascular disease ([Player & Peterson, 2011](#)), one of the costliest public health concerns. Furthermore, anxiety can occur and has been measured in a variety of settings (e.g. at the dentist; [Campos, Presoto, Martins, Domingos, & Maroco, 2013](#)).

In the mental health field, the presence and severity of anxiety symptoms are frequently measured by the Generalized Anxiety Disorder-7 scale (GAD-7; [Spitzer, Kroenke, Williams, & Löwe, 2006](#)). This is a self-report questionnaire designed to screen for severity of symptoms associated with generalized anxiety disorder ([Rutter & Brown, 2017](#)). The GAD-7 correlates highly with other measures of anxiety and it is also used in detecting the presence of many specific anxiety disorders. The GAD-7 was developed as a brief, self-report measure of anxiety, through assessment of symptoms of anxiety ([Spitzer et al., 2006](#)). For more information regarding the development, norming, and psychometric testing of the GAD-7, see [Spitzer et al. \(2006\)](#). We conducted a search on PsychINFO for published peer-reviewed articles that cited the GAD-7 since it was published in 2006. This yielded a result of 858 citations, or about 86 citations per year. The GAD-7 has been studied and validated among young adults and the elderly ([Stein et al., 2011](#); [Wild et al., 2014](#)), pregnant women ([Zhong et al., 2015](#)), diverse cultures and languages ([García-Campayo et al., 2010](#); [Sidik, Arroll, & Goodyear-Smith, 2012](#)), and a variety of physical and mental health issues in community places, such as hospitals and clinical settings ([Brown, Kroenke, Theobald, Wu, & Tu, 2010](#); [Kroenke, Spitzer, Williams, & Löwe, 2010](#)), demonstrating its versatility and utility. The measure is available in multiple languages, including, but not limited to: Portuguese ([Moreno et al., 2016](#)), Spanish ([García-Campayo et al., 2010](#)), Chinese ([Tong, An, McGonigal, Park, & Zhou, 2016](#)), German ([Löwe et al., 2008](#)), and Turkish ([Konkan, Şenormancı, Güçlü, Aydın, & Sungur, 2013](#)). Additionally, it has remained popular in comparison to other well-known tests of anxiety and is often considered the superior instrument ([Dear et al., 2011](#); [Ruiz et al., 2011](#); [Williams, 2014](#)). Furthermore, a recent study found that the GAD-7 was invariant across men and women ([Rutter & Brown, 2017](#)), indicating its usefulness across gender groups.

Research on the effectiveness of treatments that target anxiety generally utilize a pre/post-test design which can indicate statistically significant differences between a clinical sample at two time points. When large sample sizes are available, tests of statistical significance are often used in mental health research to evaluate whether or not treatments are associated with client change. Statistical significance measures how likely any differences

in outcome between treatment and control groups are real and not due to chance (Leung, 2001). Even though this is an indispensable approach to research the effectiveness of treatment, it bears some limitations. For example, Cohen (1994) demonstrated that given a large enough sample, any difference can be statistically significant even if it lacks real-world significance. Furthermore, statistical significance does not indicate whether differences that occur are meaningful. Kendall, Marrs-Garcia, Nath, and Sheldrick (1999) referred to this quality as the “convincingness of the amount of change linked to treatment” (p. 295; emphasis in original). It is important to establish that change is meaningful and not due to error in the measurement so that clinicians, researchers, and clients themselves can objectively corroborate the subjective experiences of change that occur within clinical treatments.

To assess whether or not changes are meaningful, clinicians are beginning to evaluate results with an eye for clinical significance (Kazdin, 1999). Clinical significance measures how large treatment effects are in clinical practice (Leung, 2001). Whereas the methods described above are useful for research purposes with large sample sizes, it is equally important for clinicians to have a useful tool that they can use to evaluate change on an individual, client-by-client basis. Establishing clinical significance (i.e. that a change within an individual client from Time 1 to Time 2 is real and not due to chance) is important for clinicians who want to determine and demonstrate quantitatively measurable change. Clients may also benefit from such knowledge. For example, two studies have shown that clients who received information regarding their measurement results reported higher levels of self-esteem and hope, and reported fewer symptoms compared to those who did not receive such information (Finn & Tonsager, 1992; Newman & Greenway, 1997). Kazdin (1999) also suggested clinical significance is important at the societal level, particularly in regard to issues of managed care, reimbursement, and accountability. Thus, understanding and establishing clinical significance of widely used measures may be beneficial at multiple levels of the mental healthcare system.

There are various methods for establishing meaningful clinical change (Bauer, Lambert, & Nielsen, 2004; Ferrer & Pardo, 2014; Wise, 2004). Jacobson and Truax (1991) developed what is perhaps the preferred method (Ferrer & Pardo, 2014; Wise, 2004). This involves establishing a reliable change index (RCI), which is a minimum difference between a Time 1 score and Time 2 score. This method requires that two conditions must be met in order to establish this type of change. The first criterion for doing so according to their method is the cutoff score, which refers to the lowest or highest possible score for an individual to call within a particular category. Any changes in score must move across the cutoff score, for example going from moderate to severe anxiety, in order to be considered meaningful or clinically significant change. For this study, established cutoff scores will be used. Cutoff scores will be further described in the Method section below.

The second criterion is determining that the change is statistically reliable (Jacobson & Truax, 1991). Classical test theory holds that an observed score on a measure is a combination of the true score and measurement error. In order to establish confidence that changes in scores across time that represent real changes in an individual's anxiety and are not due to measurement error, a reliable change index (RCI) must be established. The RCI helps establish that the change is not due to chance or error, but rather to real change. The following equation represents a reliable change index:

$$RCI = \frac{x_2 - x_1}{S_{diff}} \quad (1)$$

In this equation,  $x_2 - x_1$  represents an individual's change between administrations of the instrument,  $S_{\text{diff}}$  the standard error (SE) of the difference between the two scores, is defined in the following equations:

$$S_{\text{diff}} = \sqrt{2(\text{SE})^2} \quad (2)$$

$$\text{SE} = s_1\sqrt{1 - r_{xx}} \quad (3)$$

$S_{\text{diff}}$  accounts for the variation in reliability of the test instrument, and represents the standard deviation of the clinical population at intake ( $s_1$ ) and the test-retest reliability ( $r_{xx}$ ) of the instrument in a non-clinical sample. In order to solve for the formulas, it is necessary to collect data from two different samples. One is a clinical sample (in order to calculate the standard deviation for a clinical population) and the other is a community sample (in order to establish test-retest reliability in a non-clinical population). One previous study has found an RCI of the GAD-7 (Gyani, Shafran, Layard, & Clark, 2013), but the measure of reliability these researchers used was Cronbach's alpha rather than the test-retest reliability. Using the test-retest reliability provides a more accurate indicator of the instrument's reliability over time than the alpha, which has been shown to provide an overestimate of reliability and therefore an unacceptable rate of false positives (Ferrer & Pardo, 2014).

The purpose of this study is to provide clinicians and researchers with information to determine whether clients have made clinically significant change in anxiety as measured by the GAD-7. Using previously established cutoff scores and Jacobson and Truax's (1991) method, we will establish a Reliable Change Index which, when applied to two administrations of the GAD-7, will indicate if a client has experienced meaningful change. The current research is modeled after a previous paper which established the Reliable Change Index for the Revised Dyadic Adjustment Scale (Anderson et al., 2014).

## Method

### Clinical Sample

All study procedures were approved by the [redacted] Institutional Review Board, and informed consent was obtained from all individual participants included in the study. This study utilized archival data from clients seen for at least one therapy session at a university clinic in the Northeastern United States between 2008 and 2012. A total of 829 cases began therapy from March 2010 to September 2014. Clients were included in the study if they scored lower than the cutoff of 25 (i.e. in the "distressed" range) on a measure of general functioning, the Outcome Rating Scale (ORS; Miller, Duncan, Brown, Sparks, & Claud, 2003). This left 354 cases remaining. This criterion was chosen to only include clients who are experiencing clinical levels of distress. Data are present from 332 individuals after using listwise deletion to remove 22 cases that did not fully complete the GAD-7. Of the 22 deleted cases, 13 were missing all items, one was missing five items, three were missing two items, and five were missing one item. Listwise deletion was chosen because variance, which is a critical input value for this study, and would be inflated if cases with missing values were included or imputed.

The average age of participants was 25.79 ( $SD = 9.8$ ) with a range from 18 to 78. Participants were 68.7% female and were 72.9% White, 5.4% Asian, 5.1% Latino/a, 2.7% Black, and 4.5% in other racial categories, with 6.9% preferring not to answer. Most clients had at least a high school education (98%), and of those 29.6% had a Bachelor's degree or higher. There was a wide distribution of income groups. About 21% (20.8) of

clients had a household income of less than \$20,000 per year; 24% fell between \$20,000 per year and \$70,000 per year; 18.7% had an income of greater than \$70,000 per year; 32.8% either did not know their household income or preferred not to answer. Around 76% of clients were admitted for individual therapy; 16.3% for couples therapy; 4.2% for family therapy; and 3% for high-conflict co-parenting treatment.

## Community Sample

A sample was drawn from the general population in order to determine test-retest reliability of the GAD-7. A non-clinical sample must be drawn in order to test this property because presumably clients would be undergoing change while receiving therapy, which would impact test-retest reliability scores. Participants were recruited through Amazon Mechanical Turk (MTurk). MTurk is an online market where tasks (in this case a survey) are posted for respondents to complete for a specified rate. Typical tasks may include choosing between potential photographs for an advertisement, writing product descriptions, or filling out surveys and pay rates generally range from \$5 to \$7 an hour for survey work. The site is hosted by Amazon.com and all respondents have an Amazon account through which all transactions are handled. As such, each respondent is completely anonymous to the researcher who posts the “job”.

MTurk workers interested in participating viewed the job description, which included the exclusion criteria that participants must be at least 18 years old, currently residing in the United States, and not currently seeing a therapist. This final criterion was selected to capture an estimated test-retest reliability as precise as possible. In order to further determine if this sample was distinct from the clinical sample in terms of mental health distress, we also assessed these participants' level of depression using the Major Depression Inventory (MDI; Olsen, Jensen, Noerholm, Martiny, & Bech, 2003).

Of the 370 people to view the information page at the beginning of the survey, 173 participants completed the survey. Participants entered their MTurk Worker ID, which is a unique identifying number, so that their Time 2 survey could be matched to their Time 1 survey. The researchers contacted the participants via MTurk at two, three, and four weeks following the first survey to instruct them to complete the second one. Of the 173 participants who completed the GAD-7 during the first survey, 116 participants returned and completed the Time 2 survey. On average, participants filled out the Time 2 survey after 15.8 days ( $SD = 3.06$ ). A final sample of 111 remained after listwise deletion. Of those deleted, all five had only one item missing.

Participants were 47.7% female and were 79% White, 9% Asian, 5.4% Black, 1.8% multiracial, and 4.5% in other racial categories. The average age of participants was 42.7 ( $SD = 14.1$ ) with a range from 21 to 87. All participants but one had at least a high school education, with 58.5% having a Bachelor's degree or higher. There was a wide distribution of income groups. Approximately 16% (16.2) of respondents had a household income of less than \$20,000 per year; 55.8% fell between \$20,000 per year and \$70,000 per year; 27.9% had an income of greater than \$70,000 per year.

## Measures

### Generalized Anxiety Disorder-7 Scale (GAD-7)

The GAD-7 is a seven-item self-report measure of the severity of the symptoms of generalized anxiety disorder (Spitzer et al., 2006). Instructions ask: “Over the last 2 weeks, how often have you been bothered by the following problems?”. Example items include: “Feeling nervous, anxious, or on edge” and “Being so restless

that it's hard to sit still." Respondents answer on a 0 to 3 scale from "Not at all" to "Nearly every day". Scores range from 0 to 21 with higher scores indicating greater levels of anxiety. Scores of 0 to 4 indicate minimal anxiety; 5 to 9 mild anxiety; 10 to 14 moderate anxiety; and 15 to 21 severe anxiety. In other words, the scores of 5, 10, and 15 are the cutoff scores for mild, moderate, and severe anxiety, respectively (Spitzer et al., 2006). The scale has been found to have good internal reliability (Cronbach's alpha = .90; Spitzer et al., 2006). For the present study, alpha scores were also good (.89 for clinical sample; .94 for community sample at Time 1; .94 for community sample at Time 2).

### Major Depression Inventory (MDI)

The MDI is a 10-item self-report measure of the level and severity of depression (Olsen et al., 2003). Instructions provide the prompt, "How much of the time..." and respondents answer on a 0 to 5 scale from "At no time" to "All the time". Example items include "Have you lost interest in your daily activities?" and "Have you felt that life wasn't worth living?". Possible scores range from 0 to 50, with higher scores indicating greater depression. Scores from 20-24 indicate mild depression; 25-29 indicate moderate depression; and 30-50 indicate severe depression. This leads to cutoff scores of 20, 25, and 30 for mild, moderate, and severe depression, respectively (Cuijpers, Dekker, Noteboom, Smits, & Peen, 2007). This scale has been found to have good internal reliability (Cronbach's alpha = .90; Olsen et al., 2003). For the present study, alpha scores were also good (.87 for clinical sample; .95 for community sample at Time 1; .95 for community sample at Time 2).

## Results

For the GAD-7, the mean score for the clinical sample was 10.57 (range from 0 to 21,  $SD = 5.6$ ). For the community sample at Time 1, the mean score was 4.14 (range from 0 to 21,  $SD = 4.96$ ). For the MDI, the mean score for the clinical sample was 23.71 (range from 0 to 50,  $SD = 10.28$ ). For the community sample at Time 1, the mean score was 10.07 (range from 0 to 46,  $SD = 11.53$ ). As to be expected, mean scores were lower for the community sample on both measures and are in the lowest category of distress. These sample characteristics lend confidence to our assumption that the community sample is indeed "nonclinical". The means and standard deviations for both the clinical and community samples (at Time 1) are represented in Table 1.

A Pearson's correlation was computed to assess the 14-28-day test-retest reliability of the GAD-7,  $r(110) = .87$ , indicating good test-retest reliability. Jacobson and Truax's (1991) method for determining reliable change was used to determine the amount of change in score from pretest to posttest that would be statistically significant at the  $p = .05$  level for the GAD-7. We used the community sample's test-retest reliability estimate ( $r_{xx} = .87$ ), as well as the standard deviation of the clinical sample at intake ( $s_1 = 5.60$ ) as inputs for Equation 3. These values are presented in Table 1 below. Solving for Equation 3 resulted in a value of 2.02, which we used as the  $SE$  value in Equation 2. Solving for Equation 2 resulted in an  $S_{diff}$  for our sample of 2.85. Finally, we used this  $S_{diff}$  value in Equation 1 to solve for a level of change that would yield a significant value (i.e., 1.96). This resulted in an RCI of 5.59. For practical use the RCI would be rounded to 6.

Table 1

Means, Standard Deviations, Cutoff, and RCI of the GAD-7

Instrument	Clinical Sample			Community Sample			Cut-off	RCI
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>		
GAD-7	14.6	3.43	375	4.09	4.93	111	10	6

## Discussion

When evaluating client change it is important to be able to determine that measured change is real and not due to measurement error. Further, it is important to determine if change in single cases is clinically significant. To accomplish these goals, it is necessary to know how much change in score on a given instrument is a reliable change. This study established a reliable change index for a commonly used measure of anxiety. The [Jacobson and Truax \(1991\)](#) method was followed, with the RCI calculated using the means and standard deviations of both a clinical and community sample. Results indicated that an individual whose score on the GAD-7 moves across the cutoff of 10 and changes by six or more points from the first to the most recent administration can be classified as experiencing clinically significant change. Changes of at least six points that do not cross a cutoff can still be classified as either reliable improvement or deterioration (depending on the direction); however, they are not considered clinically significant if the client does not cross over from one severity classification to another. This finding is important because it offers a way in which clinicians and researchers can judge the importance of changes in scores on the GAD-7.

## Limitations

Though this study seeks to establish a standardized reliable change index, it should be noted that multiple choices impact the findings. First, the time frame that is selected for the test-retest reliability affects the RCI, because any variation in this result impacts the test-retest reliability estimate. For this study, a two-to-four-week time period was given for participants to complete the retest. This time frame was chosen to reflect the common interval between administrations of the GAD-7 in routine clinical practice. Though participants completed the second survey an average of 15.8 days after the first survey, the variability in the exact time frame in which the retest was completed and the selection of this particular frame impacts the test-retest reliability results. For example, the lower the test-retest reliability correlation, the less precise the measure is across time. Subsequently, the RCI would increase.

The unique composition of the clinical sample also impacted results, and features of the sample should be taken into consideration when interpreting results. Participant clients were seen for a minimum of one session at a university clinic. The sample represents clients with a variety of presenting problems, participating in different treatment modalities, including individual, couple, and family therapy, in a treatment-as-usual condition. While this diversity in the sample increases the external validity of the results of this research, the variability in sample also inflates the standard deviation of the GAD-7, which in turn influences the RCI. In order to preserve the variability in presenting problems and treatment modalities while minimizing the effect on standard deviation, we limited the clinical sample to those individuals who scored in the distressed range on a measure of general functioning. Given the target population for the measure is distressed individuals in a clinical setting,

we believe this was a warranted inclusion criterion. In sum, while efforts were undertaken to minimize threats to the reliability of the RCI, the sample diversity was maintained in an effort to enhance external validity.

We compare our results to those of Gyani et al. (2013), who found a cutoff score for the GAD-7 of 3.53 by using an estimate of internal consistency rather than test-retest reliability ( $\alpha = .90$  compared to our  $r_{xx} = .87$ ) and a standard deviation of a clinical sample that included only clients who scored above a cutoff for either anxiety or depression ( $SD = 4.41$  compared to our  $SD = 5.6$ ). Whereas both estimates of variability (i.e. the reliability of the measure and the standard deviation) impact the reliable change index, it may be that concerns about overestimating reliability by using a measure of internal consistency rather than the test-retest reliability may be less warranted than previously thought (Ferrer & Pardo, 2014). On the other hand, a future direction for research may be to more closely examine the impact of selection criteria for clinical samples when they are used to calculate an RCI.

Another limitation of this study concerns the generalizability of the findings. As the clients in the clinical sample were drawn from a university training facility, clients who seek services at this type of facility may differ from those who seek services as community-based agencies or private practices. Additionally, sampling persons on MTurk versus other potential community members could prove to be a limitation to this study as MTurk may attract a sample that is not representative of the overall population. The two samples are different in several ways, including age, education, income level, and therapy involvement. Sample respondents also provided data via two different mechanisms. While not anticipated to be statistically relevant, it is possible that differences in sample in some way impacted results. Furthermore, both the clinical and community samples were predominantly White. We recognize that the samples used in this study may limit the generalizability of the findings.

The researchers attempted to obtain a clinical sample that was as close to a 'treatment as usual' condition as possible. Respondents represent a variety of treatment modalities, presenting problems, and personal demographic conditions. While this strategy increased external validity by more closely matching conditions in community clinics, there were threats to internal validity posed by the heterogeneous nature of the sample. Further study is needed, with both homogeneous and heterogeneous samples, to establish the stability of the RCI.

Finally, it must be noted that the method we chose for establishing clinically significant change (Jacobson & Truax, 1991) itself is not 100% reliable. Despite offering an acceptably low rate of false positives (Ferrer & Pardo, 2014), individual change scores on the GAD-7 should still be interpreted with caution and involve taking clinical expertise and client self-report into account (Bornstein, 2017; Kazdin, 1999).

## Implications

The GAD-7 is a widely used assessment that is used to establish diagnoses during the intake process. The establishment of a reliable change index broadens the utility of this measure so that it can be reliably used over time. Cutoff scores and reliable change indices are helpful to clinicians who want to examine their client's progress in a manner that is in accordance with results-based accountability standards. These standards are useful to clinicians who want to establish the effectiveness of their treatment with various clients. Clinicians could anchor these findings in their practice by administering the GAD-7 at regular intervals and monitoring any changes. Administering measures repeatedly like this typically would come at a high cost to practicing

clinicians, as instruments cost money and/or may be difficult to obtain. However, the GAD-7 is free and easily accessible online. By using the RCI with the GAD-7, clinicians have an empirically supported measurement to assess client progress, improve treatment process, satisfy the demands of managed care, and thereby receive approval for continuing treatment.

The standards are also useful to researchers who want to explore effectiveness across models or conditions of clinical treatment. Though there are many ways to measure client change, reliable change indices can provide a degree of confidence that the reported change is not due to error in the instrument. Researchers could also anchor the current findings in their work by testing associations between clinically significant change and aspects of therapy that purport to create change. Furthermore, previous findings from research that used the GAD-7 could be enhanced, corrected, or re-verified by implementing use of the RCI developed by this study. This study also adds strength to and promotes the need for establishing RCIs for other mental health and relationship assessment devices.

Finally, it is our ultimate aim that clients and the community will benefit the most from the establishment of reliable change indices. First, clients can benefit directly by understanding that a measurement tool validates their degree of change. It can also empower them to continue progressing towards a healthier self. They also benefit indirectly through the clinicians that are able to provide better services to them. Implementation of a measurement that indicates whether or not meaningful change has occurred may more quickly improve client problems. This type of care will likely lead to positive feelings among clients that can then lead to more healthy behaviours and productivity in their community. Furthermore, if a community of health workers (e.g. physicians, therapists, psychiatrists, etc.) have similar understanding of and utilize the meaningful change for the GAD-7 then greater collaborative treatment can occur, providing clients the best possible outcome.

## Funding

This study was funded by the Department of Human Development and Family Studies at the University of Connecticut.

## Competing Interests

The authors have declared that no competing interests exist.

## Acknowledgments

The authors have no support to report.

## References

- Anderson, S. R., Tambling, R. B., Huff, S. C., Heafner, J., Johnson, L. N., & Ketring, S. A. (2014). The development of a reliable change index and cutoff for the Revised Dyadic Adjustment Scale. *Journal of Marital and Family Therapy, 40*(4), 525-534. <https://doi.org/10.1111/jmft.12095>
- Barry, J. A., Folkard, A., & Ayliffe, W. (2014). Validation of a brief questionnaire measuring positive mindset in patients with uveitis. *Psychology, Community & Health, 3*(1), 1-10. <https://doi.org/10.5964/pch.v3i1.76>
- Bauer, S., Lambert, M. J., & Nielsen, S. L. (2004). Clinical significance methods: A comparison of statistical techniques. *Journal of Personality Assessment, 82*(1), 60-70. [https://doi.org/10.1207/s15327752jpa8201\\_11](https://doi.org/10.1207/s15327752jpa8201_11)

- Bornstein, R. F. (2017). Evidence-based psychological assessment. *Journal of Personality Assessment*, *99*(4), 435-445.
- Brown, L. F., Kroenke, K., Theobald, D. E., Wu, J., & Tu, W. (2010). The association of depression and anxiety with health-related quality of life in cancer patients with depression and/or pain. *Psycho-Oncology*, *19*(7), 734-741. <https://doi.org/10.1002/pon.1627>
- Campos, J. A. D. B., Presoto, C. D., Martins, C. S., Domingos, P. A. D. S., & Maroco, J. (2013). Dental anxiety: Prevalence and evaluation of psychometric properties of a scale. *Psychology, Community & Health*, *2*, 19-27. <https://doi.org/10.5964/pch.v2i1.18>
- Carvalho, J., Marques, M. M., Ferreira, M. B., & Lima, M. L. (2016). Construct validation of the Portuguese version of the restraint scale. *Psychology, Community & Health*, *5*(2), 134-151. <https://doi.org/10.5964/pch.v5i2.170>
- Cohen, J. (1994). The earth is round ( $p < .05$ ). *The American Psychologist*, *49*(12), 997-1003. <https://doi.org/10.1037/0003-066X.49.12.997>
- Cuijpers, P., Dekker, J., Noteboom, A., Smits, N., & Peen, J. (2007). Sensitivity and specificity of the Major Depression Inventory in outpatients. *BMC Psychiatry*, *7*(1), 39. <https://doi.org/10.1186/1471-244X-7-39>
- Dear, B. F., Titov, N., Sunderland, M., McMillan, D., Anderson, T., Lorian, C., & Robinson, E. (2011). Psychometric comparison of the Generalized Anxiety Disorder Scale-7 and the Penn State Worry Questionnaire for measuring response during treatment of generalised anxiety disorder. *Cognitive Behaviour Therapy*, *40*(3), 216-227. <https://doi.org/10.1080/16506073.2011.582138>
- Dias, J. C. R., Silva, W. R., Maroco, J., & Campos, J. A. D. B. (2015). Escala de Estresse Percebido aplicada a estudantes universitárias: Estudo de validação. *Psychology, Community & Health*, *4*(1), 1-13. <https://doi.org/10.5964/pch.v4i1.90>
- Ferrer, R., & Pardo, A. (2014). Clinically meaningful change: False positives in the estimation of individual change. *Psychological Assessment*, *26*(2), 370-383. <https://doi.org/10.1037/a0035419>
- Finn, S. E., & Tonsager, M. E. (1992). Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy. *Psychological Assessment*, *4*(3), 278-287. <https://doi.org/10.1037/1040-3590.4.3.278>
- García-Campayo, J., Zamorano, E., Ruiz, M. A., Pardo, A., Pérez-Páramo, M., López-Gómez, V., . . . Rejas, J. (2010). Cultural adaptation into Spanish of the Generalized Anxiety Disorder-7 (GAD-7) scale as a screening tool. *Health and Quality of Life Outcomes*, *8*(1), 8. <https://doi.org/10.1186/1477-7525-8-8>
- Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy*, *51*(9), 597-606. <https://doi.org/10.1016/j.brat.2013.06.004>
- Heafner, J., Silva, K., Tambling, R. B., & Anderson, S. R. (2016). Client-reported-presenting problems at an MFT clinic. *The Family Journal*, *24*(2), 140-146. <https://doi.org/10.1177/1066480716628581>
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12-19. <https://doi.org/10.1037/0022-006X.59.1.12>
- Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology*, *67*(3), 332-339. <https://doi.org/10.1037/0022-006X.67.3.332>

- Kendall, P. C., Marrs-Garcia, A., Nath, S. R., & Sheldrick, R. C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology, 67*(3), 285-299. <https://doi.org/10.1037/0022-006X.67.3.285>
- Konkan, R., Şenormancı, O., Güçlü, O., Aydın, E., & Sungur, M. Z. (2013). Validity and reliability study for the Turkish adaptation of the Generalized Anxiety Disorder-7 (GAD-7) scale. *Archives of Neuropsychiatry, 50*, 53-58. <https://doi.org/10.4274/npa.y6308>
- Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2010). The patient health questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. *General Hospital Psychiatry, 32*(4), 345-359. <https://doi.org/10.1016/j.genhosppsych.2010.03.006>
- Leung, W. C. (2001). Balancing statistical and clinical significance in evaluating treatment effects. *Postgraduate Medical Journal, 77*(905), 201-204. <https://doi.org/10.1136/pmj.77.905.201>
- Losoi, H., Turunen, S., Wäljas, M., Helminen, M., Öhman, J., Julkunen, J., & Rosti-Otajärvi, E. (2013). Psychometric properties of the Finnish version of the Resilience Scale and its short version. *Psychology, Community & Health, 2*, 1-10. <https://doi.org/10.5964/pch.v2i1.40>
- Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care, 46*(3), 266-274. <https://doi.org/10.1097/MLR.0b013e318160d093>
- Marques, M. M., De Gucht, V., Gouveia, M. J. P. M., Cordeiro, A., Leal, I. P., & Maes, S. (2013). Psychometric properties of the Portuguese version of the Checklist of Individual Strength (CIS20-P). *Psychology, Community & Health, 2*, 11-18. <https://doi.org/10.5964/pch.v2i1.57>
- Miller, S. D., Duncan, B. L., Brown, J., Sparks, J. A., & Claud, D. A. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy, 2*(2), 91-100.
- Moreno, A. L., DeSousa, D. A., Souza, A. M. F. L. P., Manfro, G. G., Salum, G. A., Koller, S. H., . . . Crippa, J. A. D. S. (2016). Factor structure, reliability, and item parameters of the Brazilian-Portuguese version of the GAD-7 questionnaire. *Temas em Psicologia, 24*(1), 367-376. <https://doi.org/10.9788/TP2016.1-25>
- Newman, M. L., & Greenway, P. (1997). Therapeutic effects of providing MMPI-2 test feedback to clients at a university counseling service: A collaborative approach. *Psychological Assessment, 9*(2), 122-131. <https://doi.org/10.1037/1040-3590.9.2.122>
- Olsen, L. R., Jensen, D. V., Noerholm, V., Martiny, K., & Bech, P. (2003). The internal and external validity of the Major Depression Inventory in measuring severity of depressive states. *Psychological Medicine, 33*(2), 351-356. <https://doi.org/10.1017/S0033291702006724>
- Pimenta, F., Leal, I., & Maroco, J. (2012). The Portuguese version of the perceived control over hot flushes index: Evaluation of its psychometric properties. *Psychology, Community & Health, 1*(2), 221-231. <https://doi.org/10.5964/pch.v1i2.33>
- Player, M. S., & Peterson, L. E. (2011). Anxiety disorders, hypertension, and cardiovascular risk: A review. *International Journal of Psychiatry in Medicine, 41*(4), 365-377. <https://doi.org/10.2190/PM.41.4.f>

- Ruiz, M. A., Zamorano, E., García-Campayo, J., Pardo, A., Freire, O., & Rejas, J. (2011). Validity of the GAD-7 scale as an outcome measure of disability in patients with generalized anxiety disorders in primary care. *Journal of Affective Disorders, 128*(3), 277-286. <https://doi.org/10.1016/j.jad.2010.07.010>
- Rutter, L. A., & Brown, T. A. (2017). Psychometric properties of the Generalized Anxiety Disorder Scale-7 (GAD-7) in outpatients with anxiety and mood disorders. *Journal of Psychopathology and Behavioral Assessment, 39*(1), 140-146. <https://doi.org/10.1007/s10862-016-9571-9>
- Sidik, S. M., Arroll, B., & Goodyear-Smith, F. (2012). Validation of the GAD-7 (Malay version) among women attending a primary care clinic in Malaysia. *Journal of Primary Health Care, 4*(1), 5-11. <https://doi.org/10.1071/HC12005>
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Stein, C. H., Abraham, K. M., Bonar, E. E., Leith, J. E., Kraus, S. W., Hamill, A. C., . . . Fogo, W. R. (2011). Family ties in tough times: How young adults and their parents view the US economic crisis. *Journal of Family Psychology, 25*(3), 449-454. <https://doi.org/10.1037/a0023697>
- Tong, X., An, D., McGonigal, A., Park, S. P., & Zhou, D. (2016). Validation of the Generalized Anxiety Disorder-7 (GAD-7) among Chinese people with epilepsy. *Epilepsy Research, 120*, 31-36. <https://doi.org/10.1016/j.eplepsyres.2015.11.019>
- Wild, B., Eckl, A., Herzog, W., Niehoff, D., Lechner, S., Maatouk, I., . . . Löwe, B. (2014). Assessing generalized anxiety disorder in elderly people using the GAD-7 and GAD-2 scales: Results of a validation study. *The American Journal of Geriatric Psychiatry, 22*(10), 1029-1038. <https://doi.org/10.1016/j.jagp.2013.01.076>
- Williams, N. (2014). The GAD-7 questionnaire (Questionnaire review). *Occupational Medicine, 64*(3), 224-224. <https://doi.org/10.1093/occmed/kqt161>
- Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment, 82*(1), 50-59. [https://doi.org/10.1207/s15327752jpa8201\\_10](https://doi.org/10.1207/s15327752jpa8201_10)
- Zhong, Q. Y., Gelaye, B., Zaslavsky, A. M., Fann, J. R., Rondon, M. B., Sánchez, S. E., & Williams, M. A. (2015). Diagnostic validity of the Generalized Anxiety Disorder-7 (GAD-7) among pregnant women. *PLOS ONE, 10*(4), Article e0125096. <https://doi.org/10.1371/journal.pone.0125096>