Depression Anxiety Stress Scale-10:

A Brief Measure for Routine Psychotherapy Outcome and Progress Assessment

W. Kim Halford   &   Aaron D. J. Frost

University of Queensland   Benchmark Psychology & Griffith University

Correspondence about this article to Kim Halford, School of Psychology, University of Queensland, St. Lucia Qld. 4072 Australia. Email: K.Halford@psy.uq.edu.au

Aaron D. J. Frost, Benchmark Psychology, Level 1 Building 10 107 Miles Platting Road, Eight Mile Plains, Qld 4113 Australia. Email: aaron@benchmarkpsychology.com.au

Abstract

Routine outcome measurement and progress monitoring is well established to enhance quality assurance in clinical psychology service delivery, but is not widely used in routine care. A major barrier to more widespread implementation is the lack of public domain, brief, psychometrically sound outcome measures that easily integrate into clinical information systems. The current study assessed a brief 10-item version of the widely used Depression Anxiety Stress (DASS)-42 scale, which we called the Depression Anxiety Stress-10 (DASS-10) scale. In two clinical samples of adults (n = 1036, 445 men, 591 women; and n = 1084 493 men, 591 women), the DASS-10 had a replicable two-level factor structure, which at the lower level had two factors assessing stress-anxiety and depression, which each loaded onto a superordinate psychological distress scale. The items in the distress score discriminated between a clinical sample (n = 376) and a community sample (n =379); and were sensitive to clinical change. The measure has potential to make routine outcome measurement and progress monitoring more cost-effective to implement than existing measures, particularly when integrated with practice management software to make administration, scoring, and use easy.

Depression Anxiety Stress Scale-10:

A Brief Measure for Routine Psychotherapy Outcome and Progress Assessment

Clinicians substantially overestimate their psychotherapeutic effectiveness (Walfish et al., 2012), and are particularly poor at detecting when clients deteriorate (Hatfield et al., 2010). In adult outpatient psychological treatment about 5-10% of clients deteriorate, and between 30 and 50% show no reliable improvement during treatment (Lambert, 2010). Many experts have recommended use of routine outcome measurement as a quality assurance mechanism for psychotherapy (e.g., Boswell et al., 2015; Lambert , 2010). In addition, experts recommend routine systematic monitoring of progress session by session, with feedback to the clinician and client of progress. Progress monitoring with feedback aids early identification of patients not benefiting from therapy (Lambert, 2017). In a meta-analysis of 24 studies progress monitoring with feedback enhanced therapy outcome (Lambert et al., 2018).

Despite the strong case for routine outcome assessment and progress monitoring, only a small minority (less than 20%) of clinicians routinely use either in their practice (Boswell et al., 2015; Jensen-Doss et al., 2018). The reasons for this low uptake are complex, but a significant contributor is likely to be the lack of an easily accessed, inexpensive, and psychometrically sound measure that can be used for both outcome and progress assessment (Boswell et al., 2015; Ionita & Fitzpatrick, 2014). In the current study, we developed a brief 10-item version of the widely used, public domain measure the Depression, Anxiety and Stress Scale (Lovibond & Lovibond, 1995)[1].

The Use of Outcome and Progress Assessment in Routine Practice

Routine outcome assessment involves assessment of the change in client's scores on standardized clinical instruments before and after therapy. Outcome assessment informs

psychological services about their effectiveness, which can allow refinement of services. For example, routine outcome assessment in an outpatient clinic specializing in treatment of anxiety disorders and depression (Halford et al., 2002) identified that patients who had substance abuse problems improved less than other patients. The service then offered substance abuse treatment to patients to address the lack of therapy effectiveness for those patients. Outcome evaluation also provides information to those who fund services such as the Australian national health system called Medicare), health insurance companies, and employers who fund employee assistance programs.

Routine progress monitoring involves assessing client on standardized measures of clinical status at every session across the course of therapy, which allows clinicians to determine if a client is improving during therapy. In progress monitoring, the clinician compares a client's session-by-session adjustment trajectory with a model trajectory of clients who ultimately benefited from therapy (Lambert, 2010). If the client has a trajectory toward recovery (is "on track") then the clinician and client are reassured the approach being taken is working. If the client is off track, the clinician discusses progress with the client, and takes actions such as reformulating the goals of therapy, reviewing the therapeutic relationship, or altering the treatment approach to therapy, all with the aim of achieving a better outcome for the client (Probst et al., 2020). In addition, supervisors can use progress monitoring to guide which cases to focus upon in supervision and use supervision to enhance therapeutic outcome (O'Donovan et al., 2011).

In summary, outcome assessment requires administration of a measure before and after therapy, whereas routine progress monitoring requires session-by-session administration of a measure. A challenge with relying on before and after therapy administration to assess therapy outcome is that it is never certain at what session clients might complete treatment. This often leads to conducting post-assessments after termination of therapy, which adds to

the administrative costs of gathering data. Moreover, it often results in substantial non-completion of assessments, and the resultant missing data compromises the accuracy of estimated outcomes (Swift & Greenberg, 2012). In contrast, collecting a measure every session gathers reliable outcome data, as it always provides a measure of client status at the final session. Moreover, if a measure is given every session, that also allows for progress monitoring with feedback.

Barriers to Outcome Evaluation and Progress Monitoring

A review of available measures used to assess outcome and progress shows significant limitations of the currently available measures and technology (Overington & Ionita, 2012). Numerous measures have well-established, acceptable psychometric properties such as construct validity and sensitivity to change, but also have quiet large numbers of items for repeated use across every session. For example, the Treatment Outcome Package (TOPS) (Kraus et al., 2005) has 52 items, the Outcome Questionairre-45 (OQ-45) (Lambert, 2010) has 45 items, and the Clinical Outcomes in Routine Evaluation (CORE-OM) has 34 items (Evans et al., 2002). Measures of these lengths usually take from 10 to 20 minutes to complete (Overington & Iota, 2012), and can take much longer if clients have low levels of formal education. Measures of such length tend not to be used in a sustained way in routine practice (Miller et al., 2006), presumably in part because of the burden of repeated lengthy administration places on clients.

There are some much briefer outcome measures available, but there seems to be a trade off with such brief measures having less robust psychometrics. For example, Miller et al (2003) developed the 4-item ORS as an attempt to create a very brief version of the constructs assessed by the OQ-45. The 4-item ORS has an unacceptably low three-week test-retest reliability of $r = .49$ (Miller et al, 2003), whereas the 45-item OQ-45 has an acceptable, high three-week test-retest reliability of $r = .84$ (Lambert, 2015).   A second widely used

measure is the K10 (and a briefer variant the K6), which includes items assessing anxiety and depression. The K10 was originally developed as a brief screening measure for epidemiological mental health research (Kessler et al., 2002), but subsequently has been widely used to evaluate psychotherapy outcome (e.g., Rickwood et al., 2015). However, the factor structure of the K10 has proved not to be reliably unidimensional (Berle et al., 2010). Factor analyses variously report the K10 as having four factors (Brooks et al., 2006), two factors (Lace et al., 2019), or different factors structures in community samples (two factors) versus clinical samples (one factor) (Sunderland et al., 2012).

Another barrier to sustained use of measures in routine practice is if the measure's administration and scoring is difficult. Paper-and-pencil administered measures need to be hand scored or scanned for computer scoring. In our experience, such measures are time consuming to use. Even personal computer administration can be inconvenient as it involves special testing stations to be available, it can be hard to ensure privacy in waiting rooms, and setting aside special rooms for private administration adds to practice costs. If clients do not arrive early for sessions, clients often skip providing data. One can send online links to clients ahead of sessions, but in our experience client adherence to these requests can be low. In the ecology of busy office practice, often it is impractical for the test administration and scoring to occur and have the session start on time. Moreover, for the purpose of progress monitoring, it is highly desirable to have some graphical representation of progress to present to clients, and that needs to be generated and available in easily accessible form. Tablet or mobile phone administration of brief measures can enhance ease of administration and scoring, but the system needs to be secure and easy for clients and clinicians to use.

Additional barriers to sustained use of measures in routine practice are the cost of the measure, and lack of integration of software with the clinical management system. All widely used measures have costs for anything other than the paper-and-pencil measures, and

software for computer administration is stand alone, rather than being integrated with general practice management software. In Australia, where we conducted the current study, there is no mandated requirement to evaluate psychology services, and no schedule of fees offered under Medicare to cover the costs of such assessment. For outcome and/or progress monitoring to become routine in Australian psychology practice, the assessments need to be inexpensive to access and easy to use.

In summary, there is substantial evidence supporting the use of outcome evaluation and progress monitoring. However, existing systems of outcome and progress monitoring tend to use measures that are either too long to be practical for routine clinical use, or the shorter measures have inadequate psychometric properties. Moreover, systems of administration are expensive and not integrated with broader clinical management software. Our goal was to develop a system that used a measure of 8 to 12 items length, with a replicable and coherent factor structure. We wanted to administer the measure via computer tablet for ease of access to client and clinician; to have administration, scoring, and generation of client feedback easy; and to be able to be done in one to two minutes. We suggest routine outcome measurement must be this simple, and should be able to generate clinical progress and outcome data without additional time or cost.

Depression Anxiety and Stress Scales

We developed a brief version of the Depression, Anxiety, and Stress Scale (DASS) (Lovibond & Lovibond, 1995), aiming for it to be an easily accessed, practical system for outcome and progress monitoring. We chose the DASS for the following reasons. It currently has two versions, a 42-item and a 21-item scale (Lovibond & Lovibond, 1995), which are freely available in the public domain. Both versions have well established psychometric properties, and the DASS has been used extensively across a diverse range of populations and settings (Lovibond & Lovibond, 1995; Oei et al., 2013; Wang et al., 2016). There are

extensive normative data provided, which meant that each of the items within the DASS is well understood psychometrically, which facilitated item selection for a brief measure.

In addition to the DASS items, we added two items: one assessing suicide risk and one assessing problems with substance abuse. Some of the other progress monitoring measures include items to screen for these problems. For example, the OQ-45 has an item assessing suicidal thinking (Lambert, 2010). Our rationale was that such items might prove useful to detect some individuals who had problems not assessed directly by the DASS items, but that were of clinical importance. We did not intend that the additional items would be part of the DASS, but expected the DASS scores to show small to medium correlations with the items.

We aimed to select a subset of approximately 10 items that had a coherent and replicable factor structure, had high intra-scale reliability, discriminated between clinical and community non-clinical populations, and was sensitive to clinical change. We evaluated the use of the measure in a large clinical psychology practice across a six-year period using software integrated with the practice management software to test its sustained application in a routine care setting.

Method

Participants

There were two samples in the study, a clinical and a community sample. The clinical sample was 2120 adult clients (938 men, 1182 women), who were recruited from 2184 consecutive presentations to a large clinical psychology private outpatient practice in metropolitan Brisbane, Australia between April 2013 and January 2018. Sixty-four clinic clients (3%) did not provide data at the initial presentation, either because they declined or technical problems in entering the data that existed at the start of the study.

The clinic is located approximately 15 km from the centre of Brisbane in the busy 80 km corridor linking Brisbane to the city of Gold Coast. Mean age of the sample was 35.6

years (*SD* = 14.3). Clinicians do not conduct standardized assessment of diagnosis in the clinic. Anecdotally, therapists report the overwhelming majority of presentations were for depression or anxiety disorders, with less frequent but still common presentations of eating disorders, autism, relationship problems, and psychoses. This pattern is consistent with evidence that depression and anxiety disorders are the most prevalent mental health problems in the population (World Health Organization, 2017), and the most common presentations in generalist outpatient mental health settings in Australia and other Western countries (Jorm et al., 2017). Twenty-one psychologists (4 male and 17 female psychologists) treated the clients. The psychologists treated a mean of 101.0 clients each (*SD* = 83.9), ranging from six clients to 283 clients. All these psychologists had a minimum of a Masters level qualification in clinical psychology (6 Masters; 8 Doctorate; 5 PhD). Most clients of the clinic were eligible for rebates toward the cost of treatment through the Australian national Medicare scheme and, as required by that funding, the psychologists provided evidence supported treatments. The treatments predominantly were cognitive behaviour therapy.

Sample 2 was a community sample of 376 newlywed heterosexual couples recruited originally for a study of influences on marital adjustment, and who provided normative data for comparison with the clinical sample. Couples were recruited with the assistance of the Queensland Registrar of Marriages in three cohorts, using an identical recruitment strategy, which involving a mail out described by Halford et al. (2010). In brief, the Registrar sent a letter to all couples married in the state of Queensland, Australia during April 1999; June 2000; or August 2001. An enclosed brochure described the study, and included a reply slip and prepaid envelope for couples to provide their contact information. The researchers invited couples to register for the study. We randomly selected half the couples to use the wives' data, and the other half of the couples we used the husbands' data. to ensure we had 376 independent individuals' data. Mean age of the sample was 32.8 years (*SD* = 10.3).

Measures

The Depression Anxiety Stress Scales -10 (DASS-10) has 10 items that were selected from the original DASS-42. The DASS-42 has three scales of 14 items each that assess depression, anxiety, and stress. We selected three or four items per scale that met three criteria assessed using published data (Lovibond & Lovibond, 1995). First, all selected items had high factor loadings (greater than 0.5) on the designated factor (depression, anxiety or stress, respectively), and low cross scales loadings (less than 0.3), across four independent samples (Lovibond & Lovibond, 1995). Second, items sampled different symptoms within each scale of depression (inertia, anhedonia, devaluation of life, hopelessness), anxiety (autonomic, subjective, situational) and stress (irritability, upset, difficulty relaxing). Third, items assessed symptoms that were most sensitive to individual differences at low, moderate and high levels of each scale. These criteria select items that reflect the original scales, reflect the range of symptoms underlying each factor, and provide measurement sensitivity across the continuum of severity. All items are rated on a four-point scale of severity in the past week (0 = *did not apply to me at all*, 3 = *applied to me very much or most of the time*).

A single item was used to assess substance abuse, which was "I feel annoyed by people that criticise my drinking or drug use", and a single item to assess suicidality; "I have thoughts of ending my life". Both these items were rated on the same four-point scale of symptom severity in the past week as the DASS-10 (0 = *did not apply to me at all*, 3 = *applied to me very much or most of the time*).

Procedure

While we did not seek to evaluate the central practice management software, we describe some details to provide the context for the administration of the measures in routine community practice. The programmer built the software with Filemaker Pro, which is a relational database that has a series of tables summarizing clinical practice data (e.g., client

contact details, referral contact details, client progress notes, client score on the DASS-10 at

each session, billing, and payment details), with embedded links between the tables. The

linked tables allow generation of a range of reports on clinical practice data (e.g., all clients

referred by a particular referrer, all clients who present with a DASS score of high severity).

Prior to each therapy session, clinic reception staff handed clients a tablet device that

had Filemaker Go installed, which is a mobile device version of Filemaker Pro. A

receptionist generated a Quick Response (QR) code on the reception computer, the tablet

scans the QR code and that opens the DASS for client completion on the tablet. The QR code

tags the answers with client identification, date and session number. Once the client

completes the questionnaire the data are automatically downloaded to the clinic central

database, and instantly are accessible by the therapist on their computer in their office.

Importantly, clinicians can access a graph showing the trajectory of change across sessions.

The trajectory of change was not the focus of the current study, but clinicians used it in

reviewing therapy progress with clients.

In the first session, clients are presented with an informed consent form on the tablet,

which states that the primary purpose of routine data collection is for therapy, and that de-

identified data can also be used for research evaluating clinical outcomes. The clients have

the option to withhold agreement of collection of data for one or both of these purposes, and

informed they are able to access therapy irrespective of their provision of data. Conduct of

the research complied with the ethical standards of the Helsinki Declaration of 1975, as

revised in 2008, and Griffith University Human Research and Ethics committee (Approval

number: 2018/173) approved the conduct of the study.

The community sample of 379 newlywed couples who volunteered for the study were

contacted by telephone, and each partner was individually sent a questionnaire booklet and a

prepaid return envelope. Couples who had not returned their booklets within seven days were

recontacted with up to a further two telephone calls to prompt return. Conduct of the research complied with the ethical standards of the Helsinki Declaration of 1975, as revised in 2008, and Griffith University Human Research and Ethics committee (Approval number: HLS/26/99/HRC) approved the conduct of the research.

Statistical Analysis

Approximately 50% of the clinical sample were randomly allocated into each of two samples to form the development sample (n =1036, 445 males and 591 females), and the replication sample (n = 1084, 493 males and 591 females). There was less than 0.5% of missing data in each of the two samples, and we used maximum likelihood estimation to impute the very small number of missing values. We conducted an exploratory factor analysis of client scores at presentation in the development sample of the 10 items of the DASS-10 using maximum likelihood extraction and oblimimal rotation. That analysis yielded two highly correlated factors Anxiety-Stress and Depression scales (see Results section for details). It seemed likely that there was a two-level factor structure, with the Anxiety-Stress and Depression scales loading onto a higher order factor. To test that possibility, we ran a Confirmatory Factor Analysis (CFA) in AMOS with the development sample. As is conventional in scale development, we then tested the replicability of this structure with a second CFA the structure in the replication sample.

In order to test the discriminant validity of the DASS-10, we compared the first 376 consecutive clinical presentations from the clinical sample with the community sample, which gave equal sample sizes. We used a simple t-test comparison of means to test the difference. We then assessed the variability of clinical change and the clinical significance of change within that clinical group. Following Jacobson and Truax (1991), we calculated the reliable change index (RCI) as the minimum scale point change that indicted at $p < .05$ that such a change was unlikely to be attributable to measurement error, which is calculated as

1.96 times the standard error of measurement of the change score. We classified any client with change from pre-to post-therapy equal to or greater than the RCI to have reliable change.

Clinically significant change occurs if the client showed RCI improvement, and at the end of therapy was no longer in the clinically distressed range (Jacobson & Truax, 1991). The cut off to define clinically significant change is the point where the normal curves of the community and clinical sample intersect, scores equal to or less than this distress score, mean the person is equally likely to be drawn from the community as the clinical sample, i.e. the clinical client is no longer distinguishable from a community member. The clinical cut off can be calculated using the following formula (Jacobson & Truax, 1991).

$$\text{Clinical cut off} = ((SD_{Co} * M_{Cl}) + (SD_{Cl} * M_{Co})) / (SD_{Co} * SD_{Cl})$$

SD_{Co} – the standard deviation of the community sample, and SD_{Cl} – the standard deviation of the clinical sample. M_{Co} = mean of the community sample and, M_{Cl} = mean of the clinical sample.

Results

Construct validity

The KMO was .911, showing the covariance structure was appropriate for factor analysis. The analysis extracted two factors with Eigenvalues that exceeded one, which accounted for 50.2% and 11.2% of the variance, respectively. Table 1 presents the item factor loadings. Five items loaded onto the first factor at greater than 0.5, and one item loaded at 0.42, the highest loading items related to anxiety or stress, and hence we named the first factor Anxiety-Stress. Items in the Anxiety-Stress scale had high internal consistency, α =.83. The remaining four items loaded onto the second factor and all relate to depression, so we labelled that factor Depression. Items in the Depression scale also had high internal

consistency, α =.85. There were few cross loadings between the two factors, and none that exceeded an absolute value of 0.20, but the factors correlated highly, r =.69.

As the analysis yielded two highly correlated factors Anxiety-Stress and Depression scales, it seemed likely that there was a two-level factor structure, with the Anxiety-Stress and Depression scales loading onto a higher order factor. In a CFA testing the two-level model shown in Figure 1, the model fit was acceptable, $\chi^2(34) = 218.832$, $p < .001$, CFI = .959, RMSEA = .072. As shown in Figure 1, all items loaded as expected onto the first level factors of Anxiety-Stress and Depression, which each loaded onto a higher order factor we labelled as Distress. All coefficients on the Distress scale were large and significantly different to zero at $p < .01$, and the internal reliability of Distress scale was high, $\alpha = .89$

As is conventional in scale development, we then tested the replicability of the two-level structure in the second (replication) sample. Again, the model fit was acceptable, $\chi^2(34) = 230.873$, $p < .001$, CFI = .956, RMSEA = .073. Figure 1 shows the item factor loadings. Again, all items loaded as expected with factor loadings all significant and exceeding 0.5, and loadings were similar in the development and replication samples. Moreover, the six items in the Anxiety-Stress scale showed high internal consistency in the replication sample, α =.83; as did the four items in the Depression scale, α =.85, and the 10 items in the Distress summary scale, $\alpha = .88$.

Table 2 presents the means, standard deviations and correlations between the DASS-10 and two extra items that served as validity measures: suicidal thoughts and others' expressed concerns about the respondent's substance abuse. There was a large correlation between the DASS-Anxiety scale and the DASS-Depression scale, and the correlations of each scale with the superordinate DASS Distress scale were large, all $r > .86$. Mean scores for reported concerns by others about drug abuse and suicidal thoughts were low. The modal response was to rate 0 - *did not apply to me at all* for reports of suicidal thoughts (67% and 68% of the

development sample and replication samples, respectively), and other's concern about substance use (81% of each of the development and replication samples). Suicidal thoughts had a large correlation with DASS Distress, other's concerns about substance use showed a small to medium correlation with DASS Distress.

Discriminant Validity

As expected, the clinical subsample mean at presentation on the DASS-10 ($M = 12.90$, $SD = 6.76$) was much higher, t (753) = 25.735, $p < .001$, $d = 1.87$, than the community sample ($M = 3.01$, $SD = 3.15$), $d = 1.88$ 95% CI [1.67- 2.09]. Applying Jacobson and Truax's (1991) formula, the cut off to define a clinical case is a DASS-10 score of seven or more. In the clinical sample 312/376 (83%) clients met criterion for being clinically distressed, and in the community sample it was 48/376 (13%). Thus, there was very limited overlap between the clinical and community samples. The community sample cases likely reflect the base rates of clinical distress in the community.

Sensitivity to Change

As the aim with the new measure was to evaluate outcomes and progress, it needed to be sensitive to change. We evaluated change from the initial presentation to the final session in n = 376 consecutive clinical presentations, who received a mean of 10.6 sessions ($SD = 13.0$). Mean change across all clients was 4.14 points ($SD = 5.83$), which is a medium to large effect size, $d = 0.61$ 95% CI [0.32 - 0.90]. Among the 312/376 clients presenting who met the clinical case criterion, the mean change was 5.01 points ($SD = 5.83$), $d = 0.74$ CI [0.46 - 1.02], which is a large effect size.

Applying Jacobson and Truax (1991) procedure, the reliable change index was five points (based on the internal reliability of $\alpha = 0.88$, cut off = 4.28 points). Table 3 presents the reliable and clinical significance of change. As shown, across the whole sample, 173/376 (46%) clients improved, 181 (48%) showed no reliable change, and 22 (6%) deteriorated. The

DASS-10 seemed more appropriate to measure change in clients who present meeting the

clinical cut-off for caseness (i.e., DASS-10 $\geq$ 7 at presentation), and for them 170/312 (54%)

improved, 127 (41%) showed no reliable change, and 15 (5%) deteriorated. Of the 312 clients

meeting caseness, 93 (30%) showed both reliable change and no longer met criteria for

caseness at the end of therapy. In brief, about one third of clinical cases recovered, another

20+ % improved, but more than 40% did not improve reliably.

To further evaluate sensitivity to change we classified client presentations into three

categories: mild (not meeting criteria for clinical caseness), moderate (7 < DASS-10 < 13,

which met caseness criteria but was within 1 SD of clinical cut off), severe (DASS-10 $\geq$ 13,

was more than 1 SD above the clinical cut off). There were 64 clients (17%) who had mild

presentations, 112 (30%) who were moderate, and 200 (53%) who were severe. Mean

number of sessions differed between clients with mild, non-clinical presentation (M = 7.25,

SD = 6.80) and clients with clinical moderate presentation (M = 10.4, SD =14.0), and clients

with severe presentation (M = 11.85, SD = 13.7), F (2, 236) = 3.099 p = .046 η^2 = 0.016, a

small effect size. Post hoc simple comparison of means showed moderate presentations did

not differ from severe presentations in mean number of sessions attended, t = 0.947, p = .344;

clients with severe presentations had more sessions than clients with mild, nonclinical

presentations, t = 2.479, p = .014; but moderate presentations did not differ significantly from

mild, non-clinical presentations, t = 1.559, p = .120.

Within those clients who met the clinical case criterion at presentation, the clients with

severe presentations were more likely to show reliable change (126/200, 63%) than those

with moderate presentations (44/112, 39%), $\chi^2(2)$ = 16.497 p < .001. However, the chance of

meeting the criteria for clinical recovery (DASS reliably improves and post assessment is less

than clinical cut off (seven) at completion of therapy) was lower for the severe presentations

(49/200, 25%) than for the moderate presentations (44/112, 39%), $\chi^2(1)$ = 7.502, p = .006. It

is noteworthy that every client with moderate presentation who showed reliable improvement also met the clinical recovery criteria.

Discussion

The aim of the study was to develop a brief measure of psychological distress to measure psychotherapy outcome, and facilitate therapy progress monitoring. We found a 10-item short form of the DASS scale yielded two internally consistent, discriminable but highly correlated factors of Anxiety-Stress and Depression. These two factors together accounted for over 60% of the total variance, and each loaded onto a superordinate factor of Distress. We replicated this two-level factor structure in a second sample. While using the separate Anxiety-Stress and Depression scales accounts for a little more of the total variance, using a single superordinate Distress score is easy, captures over half of the variance by itself, and provides similar association with related outcomes of drug use and suicidal thoughts. Therefore, for routine clinical use, we think a single Distress score is most useful.

The current study provides evidence of the psychometrics of the Distress Score of the DASS-10. It showed high internal reliability, $\alpha = .89$, and scores clearly discriminated between a clinical and a community sample. It was sensitive to change, particularly among clients who presented with initial DASS-10 scores in the clinical range. Replication is needed across more diverse samples, but these data suggest the DASS-10 is a potentially useful measure for use in clinical practice, and is a practical to use as both an outcome assessment and for monitoring progress.

The large practice setting that was the site of the current study used the DASS-10 routinely for six years, with close to 100% completion of measures. Aside from one or two individuals who declined to provide data at all, the only times assessments were missed was when clients arrived late for sessions and then declined to provide data on that occasion. The clinic uses data collected to provide feedback to therapists and clients on therapy progress, to

guide supervision to enhance therapists' effectiveness, and to include in reports about therapy to referral sources. In addition, the practice posts summary outcome (de-identified for both clients and therapists) on its web site, which provides public accountability https://www.benchmarkpsychology.com.au/about-us/outcome/.

In the current study we replicated that depression was a distinguishable factor even though the DASS-10 has only four items assessing depression. However, we found in the DASS-10 anxiety and stress (each assessed by three items) loaded onto a single factor, rather than being two distinguishable factors, as has been found for the DASS-21 or DASS-42 (Lovibond & Lovibond, 1995). The DASS-42 shows high correlation between anxiety and stress $r = .65$, and the substantial shared variance is likely to account for the lack of discriminability between anxiety and stress in the short DASS-10 form (Lovibond & Lovibond, 1995). Moreover, with the DASS-42 and DASS-21 there is a large correlation between the depression, anxiety and stress scales across many studies (e.g., Oei et al. 2013; Wang et al., 2016). The current finding of a higher order factor of distress that reflects shared variance across the scales is consistent with the previously reported correlations between factors. Hence, we think for most clinical purposes, the DASS-10 total distress score is likely to be of the most utility for outcome assessment or progress monitoring. This is not to say, that the depression or anxiety-stress subscales are not interpretable. If there are specific hypotheses about changes in anxiety-stress or depression, then using those subscales could be useful.

Limitations

Seventeen percent of client presentations did not meet the DASS-10 criteria for clinical distress. Anecdotally, it was our impression that the DASS-10 was not appropriate for assessing clients presenting with psychoses or autism spectrum disorders, as often these clients did not score in the clinical range on DASS items at presentation. Moreover, it is well

established that while some clients with couple relationship distress have elevated stress, anxiety or depression, perhaps as many as half of clients presenting with relationship distress are not in the clinical range on any of these symptom clusters (Halford & Pepping, 2019). There were clients in the current sample who presented with relationship distress, and for at least some of these clients the DASS-10 is likely not be an appropriate outcome measure. The Couple Satisfaction Index (Funk & Rogge, 2007) has well established reliability and validity, the 4-item and 16-item versions of the CSI are in the public domain, and have proved sensitive to change resulting from couple therapy (Halford et al., 2017). It is unlikely that any single measure is appropriate for use with all clients in a practice with diverse client presentations, but the DASS-10 does seem appropriate for the vast majority of client presentations in community-based psychology practices.

We conducted the current study in one large practice located in the southern suburbs of Brisbane, Australia. Future research should seek to replicate the psychometrics of the DASS-10, particularly in culturally and linguistically more diverse samples. The mean age of our clinical and community samples was mid-30s, which is similar to that reported for many community trials of psychotherapy (Cuipers et al., 2008), but generalizability of current findings to younger and older clients needs to be tested rather than assumed. The current study established the replicability of the factor structure and internal consistency of the times in the DASS-10. Future work needs to assess additional psychometrics, for example assessing test-retest reliability and convergent validity. Also, it would be useful to evaluate the utility of the DASS-10 in assessing change with different diagnostic presentations

There is a need to replicate the successful implementation of the administration system in other clinical settings. The owner of the practice has a long-standing commitment to routine progress monitoring and outcome assessment, and all psychologists working in the practice agree to use of the system as a condition of employment. There are many complex

influences on effective implementation of innovations in routine practice, but strong support for the innovation from key leaders in an organization is highly influential (Weaver et al., 2015).

Conclusions

The current analysis provides preliminary evidence for the successful use of the DASS-10 in routine psychological practice for outcome assessment and progress monitoring. Building the DASS-10 administration into clinical outcome software made the system easy to use, and easy access to progress and outcome data has the potential to not only improve the quality of care for clients, but also to gather ongoing evidence of the effectiveness of services in community practice. Despite 15 years of Medicare funding of psychologists to provide therapy, there is limited evidence of effectiveness in Medicare funded routine delivery of care. The paucity of effectiveness evidence remains an ongoing threat to the future of funding of psychological care in this country, particularly in the context of the COVID related economic recession and large government debt.

References

Berle, D., Starcevic, V., Milicevic, D., Moses, K., Hannan, A., Sammut, P., & Brakoulias, V. (2010). The factor structure of the Kessler-10 questionnaire in a treatment-seeking sample. *The Journal of Nervous and Mental Disease, 198*(9), 660–664. https://doi.org/10.1097/NMD.0b013e3181ef1f16.

Boswell, J. F., Kraus, D. R., Miller, S. D. & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research, 25* (1), 6-19. https://doi.org/10.1080/10503307.2013.817696

Brooks, R. T., Beard, J., & Steel, Z. (2006). Factor structure and interpretation of the K10. *Psychological Assessment, 18*(1), 62–70. https://doi.org/10.1037/1040-3590.18.1.62.

Cuipers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology, 76*(6), 909-922. http://doi.org/10.1037/a0013075

Evans, C., Connell, J., Barkham, M., & Margison, F. (2002). Towards a standardized brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry, 180*(1) 51-50. https://doi.org/10.1192/bjp.180.1.51

Funk, J., & Rogge, R. (2007). Testing the ruler with item response theory: Increasing precision of measurement for relationship satisfaction with the Couples Satisfaction Index. *Journal of Family Psychology, 21(4)*, 572-583. https://doi.org/10.1037/0893-3200.21.4.572

Halford, W. K., Bernoth-Doolan, S., & Eadie, K. (2002). Schemata as moderators of clinical effectiveness of a comprehensive cognitive behavioural program for patients with depression or anxiety disorders. *Behavior Modification, 26*, 571-593. https://doi.org/10.1177/014544502236651

Halford, W. K., Farrugia, C., Lizzio, A., & Wilson, K. L. (2010). Relationship aggression, violence and self-regulation in Australian newlywed couples. *Australian Journal of Psychology, 62,* 82-92. https://doi.org/10.1080/00049530902804169

Halford, W. K. & Pepping, C. A. (2019). What every therapist needs to know about couple therapy. *Behaviour Change, 36*(3) 121-142. https://doi.org/10.1017/bec.2019.12

Halford, W. K., Rahimullah, R., Wilson, K. L., Occhipinti, S., Busby, D., & Larson, J. (2017). Four year effects of couple relationship education on low and high satisfaction couples: A randomized clinical trial. *Journal of Consulting and Clinical Psychology, 85* (5), 495-507. https://doi.org/10.1037/ccp0000181.

Hatfield, D., McCullough, L., Frantz, S. H., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology and Psychotherapy, 17,* 25–32. https://doi.org/10.1002/cpp.656

Ionita, G., & Fitzpatrick, M. (2014). Bringing science to clinical practice: A Canadian survey of psychological practice and usage of progress monitoring measures. *Canadian Psychology, 55*(3), 187-196. http://dx.doi.org/10.1037/a0037355

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

Jensen-Doss, A., Haimes, E. M. B., Smith, A. M., Lyon, A. R. , Lewis, C. C., Stanick, C. F. & Hawley, K. M. (2018). Monitoring treatment progress and providing feedback is viewed favourably but rarely used in practice. *Administration and Policy in Mental Health and Mental Health Services Research*, *45*, 48-61. https://doi.org/10.1007/210488-016-0763-0

Jorm, A. F., Patten, S. B., Brugha, T. S., & Motjabati, R. (2017). Has increased provision of

treatment reduced the prevalence of common mental disorders? Review of the evidence

from four countries. *World Psychiatry, 16* (1), 90-99.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5269479/

Kessler, R. C., Andrews, G., Colpe, L. J., & Hiripi, E. (2002) Short screening scales to

monitor population prevalence and trends in non-specific psychological distress.

Psychological Medicine, 32(6), 959-976.

https://doi.org/10.1017/SO0033291702006074

Kraus, D. R. Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health

treatment outcome and assessment tool designed for naturalistic settings: The treatment

Outcome Package. *Journal of Clinical Psychology, 61*(3), 285-314.

https://doi.org/10.1002/JCLP20084

Lace, J. W., Greif, T. R., McGrath, A., Grant, A. F., Merz, Z., C., Teague, C. L., & Handal, J.

(2019). Investigating the factor structure of the K10 and identifying cutoff scores

denoting nonspecific psychological distress and the need for treatment. *Mental Health

and Prevention, 13*, 100-106. https://doi.org/10.1016/j.mhp.2019.01.008

Lambert, M. J. (2010) *Prevention of treatment failure: The use of measuring, monitoring, and

feedback in clinical practice*. American Psychological Association.

Lambert, M. J. (2017). Maximizing psychotherapy outcome beyond evidence-based

medicine. *Psychotherapy and Psychosomatics, 86*(2), 80-89.

https://doi.org/10.1159/000455170

Lambert, M. J., Whiplle, J. L., & Kleinstäuber, M. (2018) Collecting and delivering progress

feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy, 55*(4), 520-

537. http://dx.doi.org/10.1037/pst0000167

Medicare Benefits Schedule item 2715. (Undated).

http://www9.health.gov.au/mbs/fullDisplay.cfm?type=item&q=2715&qt=ItemID

Miller, S. D., Duncan, B. L., Brown, J., Sparks, J. A., & Chalk, M. B. (2006). Using formal
client feedback to improve retention and outcome: Making ongoing, real-time
assessment feasible. *Journal of Brief Therapy, 5* (1) 5-22.

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,* 91–100.

Oei, T. P. S., Sawang, S., Goh, Y. W., & Mukhtar, F. (2013) Using the Depression Anxiety,
Stress Scale 21 (DASS-21) across cultures. *International Journal of Psychotherapy, 4*8
(6), 1018-1029. https://doi.org/10.1080/00207594.2012.755535

O'Donovan A., Halford, W. K., & Walters, B. (2011). Toward best practice of clinical
psychology supervision. *Australian Psychologist, 46,* 101-112.
https://doi.org/10.1111/j.1742-9544.2011.00033.x

Overington, L., & Ionita, G. (2012). Progress monitoring measures: A brief guide. *Canadian
Psychology, 53* (2), 82-92. https://doi.org/10.1037/a0028017

Probst, T., Klienstauber, M., Lambert, M. J., Tritt, K., Pieh, C., Loew, T. H., Dahlbender, R.
W., & Delgadillo, J. (2020). Why are some cases not on track? An item analysis of the
assessment for signal cases during inpatient psychotherapy. *Clinical Psychology and
Psychotherapy, 27,* 559-566. https://doi.org/10.1002/cpp.2441

Rickwood, D. J., Marzer, K.R., Telford, N. R., Parker, A. G., Tanti, C. J., & McGorry, P. D.
(2015). Changes in psychological distress and psychosocial functioning in young
people accessing headspace centers for mental health problems. *Medical Journal of
Australia, 202* (10) 537-542. https://doi.org/10.5694/mja14.01696

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A

meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(4), 547-559.

https://doi.org/10.1037/a0028226

Sunderland, M., Mahoney, A., & Andrews, G. (2012). Investigating the factor structure of the

Kessler psychological distress scale in community and clinical samples of the

Australian population. *Journal of Psychopathology and Behavioral Assessment, 34*(2),

253–259. https://doi.org/10.1007/s10862-012-9276-7.

Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-

assessment bias in mental health providers. *Psychological Reports, 110,* 639–644.

http://dx.doi.org/10.2466/02.07.17.PR0.110.2.639-644

Wang, K., Shi, H.-S., Geng, F.-L., Zou, L.-Q., Tan, S.-P., Wang, Y., Neumann, D. L., Shum,

D. H. K., & Chan, R. C. K. (2016). Cross-cultural validation of the Depression Anxiety

Stress Scale–21 in China. *Psychological Assessment, 28*(5), e88 -

e100. https://doi.org/10.1037/pas0000207

World Health Organization (2017). *Depression and other common mental disorders: Global

health estimates*. Geneva: World Health Organization.

https://www.who.int/mental_health/management/depression/prevalence_global_health_

estimates/en/

Footnotes

[1]We thank Peter Lovibond and the late Syd Lovibond for putting the measure into the public domain, and Peter for supporting our attempt to develop a brief measure suitable for outcome and progress monitoring in routine care.

Table 1.

Item Factor Loadings of the Depression, Anxiety Stress Scale-10 in the Development Sample.

Number of factors in solution	Two factors		One factor
Item	Stress-Anxiety	Depression	Distress
1. I felt I was close to panic	.87	.05	.73
2. I found it difficult to work up the initiative to do things	.16	-.59	.69
3. I felt down hearted and blue	.01	-.75	.69
4. I was intolerant of anything that kept me from getting on with what I was doing.	.42	-.19	.57
5. I felt that I had nothing to look forward to.	.00	-.79	.70
6. I felt scared without any good reason	.77	.03	.67
7. Tended to over react to situations	.46	-.20	.62
8. I was worried about situations in which I might make a fool of myself	.77	.09	.63
9. I found it difficult to relax	.52	-.18	.65
10. I couldn't seem to experience any positive feelings at all.	-.04	-.84	.71

Large item loading of absolute value > 0.50 are underlined.

Table 2

Means, Standard Deviation and Correlation of DASS-10 Factor Scores, Substance Use Comments, and Suicide Contemplation

Scale	No. of items	Mean[1]	Standard[1] Deviation	DASS Anxiety	DASS Depression	DASS Distress	Substance Use	Suicidal Thoughts
DASS Anxiety	6	7.67, 7.53	4.36, 4.22		.64*	.94*	.19*	.38*
DASS Depression	4	5.51, 5.44	3.14, 3.18	.60*		.87*	.17*	.50*
DASS Distress	10	13.13, 12.98	6.80, 6.64	.92*	.86*		.20*	.47*
Substance Use	1	0.27, 0.29	0.65, 0.72	.23*	.21*	.25*		.17*
Suicidal Thoughts	1	0.44, 0.43	0.72, 0.72	.37*	.51*	.48*	.25*	

[1]Means and standard deviations are presented for development sample, then replication sample; Correlations for development sample above the diagonal, and for replication sample below the diagonal;* $p < 01$.

Table 3

Reliable and Clinically Significant Change across Whole Sample and Stratified by Severity at Presentation

Sample	N (% of whole sample)	Mean (SD) number of sessions	Number and (% within sample or subsample)			
			Deteriorated	No change	Improved	Recovered
Whole sample[1]	376 (100%)	10.63 (12.98)	22 (6%)	181 (48%)	173 (46%)	---
Clinically distressed at presentation	312 (83%)	11.33 (13.82)	15 (5%)	127 (41%)	77 (25%)	93 (30%)
Mild (non-clinical) distress[1]	64 (17%)	7.25 (6.80)	7 (11%)	54 (84%)	3 (5%)	---
Moderate clinical distress	112 (30%)	10.4 (14.02)	8 (7%)	60 (54%)	0 (0%)	44 (39%)
Severe clinical distress	200 (53%)	11.9 (13.71)	7 (45)	67 (34%)	77 (39%)	49 (25%)

[1]By definition the mild non-clinical sample present with DASS-10 scores in the non-clinical range and hence cannot meet criteria for recovery, and so we did not estimate recovery or the whole sample (which includes some mild presentations) or the mild presentation cases.

Figure 1. Confirmatory Factor Analysis Standardized Coefficients for Development and

Replication Samples, Respectively.