International Journal of Psychology

International Journal of Psychology, 2023
DOI: 10.1002/ijop.12907

Evaluating depression anxiety and stress assessment before and during the COVID-19 pandemic using generalisability theory

Scott J. Lightburn[1], Steven D. Pratscher[2,3], B. Ann Bettencourt[3], Jayne M. Hartstone[1], and Oleg N. Medvedev[1]

[1]School of Psychology, University of Waikato, Hamilton, New Zealand
[2]Pain Research and Intervention Center of Excellence, University of Florida, Gainesville, FL, USA
[3]Department of Psychological Sciences, University of Missouri, Columbia, MO, USA

The current study investigated the assessment of depression, anxiety, and stress during normal and COVID-19 pandemic conditions. Generalisability theory (G-theory) was applied to examine stable and dynamic aspects of psychological distress and the overall reliability of the Depression, Anxiety and Stress Scales (DASS-21), using data from two independent samples collected on three occasions with 2- to 4-week intervals. The US data ($n = 115$) was collected before the COVID-19 pandemic, and the New Zealand (NZ) data ($n = 114$) was obtained during the pandemic. The total DASS-21 demonstrated excellent reliability in measuring enduring symptoms of psychological distress ($G = .94–.96$) across both samples. While all DASS-21 subscales demonstrated good reliability with the pre-pandemic US sample, the subscales' reliability was below an acceptable level for the NZ sample. Findings from this study indicate that the overall psychological distress is enduring and can be reliably measured by the DASS-21 across different conditions and populations, while shifts across depression, anxiety and stress levels are likely during emergency and uncertainty, as seen in the COVID-19 pandemic.

Keywords: Depression; Anxiety and stress scale; Generalisability theory; Reliability; State and trait; Psychometrics.

According to the World Health Organization (2020), the coronavirus pandemic has impacted mental health services worldwide and will likely continue to do so over the next few years. Factors such as financial insecurity, restrictions to daily routines and disruption to mental health services along with many other environmental and social challenges are expected to further exacerbate already existing mental health issues (Alzueta et al., 2021; World Health Organization, 2020). Although research on COVID-19 varies from country to country, there is a growing consensus that as stress, loneliness and COVID-19 related worry increase, so will the prevalence and severity of mental health problems such as depression and anxiety (Kwong et al., 2021).

An expected increase in these negative psychological issues highlights the need for robust and accurate screening to identify potentially vulnerable individuals and provide targeted interventions (Alzueta et al., 2021). Additionally, levels of depression, anxiety and stress are likely to shift over time, as stress is often related to the development of anxiety and depression, and anxiety often precedes depression (Zhang et al., 2020). Therefore, it is emphasised that mental health services use appropriate psychometric tools that accurately distinguish between dynamic and enduring symptoms to prevent potential misdiagnoses, overdiagnoses and inaccurate evaluation of associated risk factors. These potential risks further highlight the importance of reliable and valid assessment of

Correspondence should be addressed to Oleg N. Medvedev, School of Psychology, Faculty of Arts and Social Sciences, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand. (E-mail: oleg.medvedev@waikato.ac.nz).

The authors would like to thank all participants who participated in this study. Open access publishing facilitated by The University of Waikato, as part of the Wiley - The University of Waikato agreement via the Council of Australian University Librarians.

There was no funding available for this study.

S. J. Lightburn designed the study, conducted statistical analyses and wrote the manuscript. S. D. Pratscher collaborated with data collection and edited the manuscript. B. A. Bettencourt collaborated with data collection and study design and edited the manuscript. J. M. Hartstone collected data and edited the manuscript. O. N. Medvedev supervised the study and data analyses and edited the manuscript.

distress symptoms in the general population by using assessment instruments specifically designed for this purpose (Kircanski et al., 2016; Medvedev et al., 2017).

The Depression Anxiety and Stress Scale (DASS) was developed through a series of factor analyses aimed at assessing symptoms of depression and anxiety; however, exploratory factor analysis revealed a third dimension resembling "stress" (e.g., irritability, agitation, nervous tension and low frustration tolerance; Lovibond & Lovibond, 1995). Thus, the scale was reconceptualised to measure depression, anxiety and stress with 42 items (DASS-42) and was later reduced to the more commonly used 21-item short-form scale (DASS-21; Lovibond & Lovibond, 1995). To date, there is no other psychometric scale that measures depression, anxiety and stress simultaneously, and is suitable for both general and clinical populations. The DASS-21 is among the most widely used and applicable self-report measures of affective symptoms, which is reflected in the number of Google Scholar citations over 13,000 to date (10 April 2023).

Although much research has examined the psychometric properties of the DASS, there are limitations to the methods of these evaluations. Studies assessing the DASS-21 propose that its three subscales have adequate construct validity, and although each of the three factors possesses distinct characteristics that differentiate themselves from one another, they share the common aspect of psychological distress (Lee, 2019; Medvedev et al., 2018). Further studies assessing the reliability of the DASS-21 generally reported good reliability, with Cronbach's alpha for the Depression, Anxiety and Stress subscales ranging from .81 to .94 (Osman et al., 2012). However, classical test theory (CTT) methods do not account for variability across scale items and the associated interactions between person, item and occasion. Similarly, methodological techniques, such as intraclass correlations coefficient (ICC) that assess temporal reliability have limited accuracy and do not control for variability across individual items and interactions between person, item and occasion (Bloch & Norman, 2012; Medvedev et al., 2017). Despite the frequent use of the DASS-21, its psychometric properties were mostly investigated using traditional psychometric methods such as CTT.

Recently, Rasch methodology was used to examine the psychometric properties of the DASS-21 and demonstrated robust reliability and internal structural validity in all subscales and the overall scale as a measure of global psychological distress (Medvedev et al., 2018). The same study developed ordinal to interval conversion tables for the DASS-21 to covert ordinal raw scores in the interval-level data to increase measurement precision. However, there were no investigations into whether the DASS-21 measures more enduring or more dynamic aspects of affective conditions. Additionally, generalisability of the DASS-21's assessment scores across samples and occasions were not examined using appropriate methodology. Therefore, further research using more appropriate methodology such as generalisability theory (G-theory; Cronbach et al., 1963; Paterson et al., 2018) is required to clearly distinguish between enduring and dynamic aspects of the DASS-21 and its subscales and to evaluate the overall reliability and generalisability of its assessment scores.

G-theory

G-theory has been increasingly applied to distinguish between state (dynamic aspect) and trait (enduring aspect) in psychometric measures (Medvedev et al., 2017). Developed by Cronbach et al. (1963), G-theory provides a more thorough and advanced statistical method for establishing the overall reliability and generalisability of assessment scores by examining sources of measurement error and their unique contributions to the overall assessment scores (Bloch & Norman, 2012). Compared to CTT, which only considers the error of measurement as a single factor, G-theory allows for a more in-depth and comprehensive way of examining both measures and constructs to being measured. Moreover, G-theory has already been used to examine psychometric scales measuring childhood depression and perceived stress (Miller et al., 2021; Paterson et al., 2018). For example, by utilising G-theory, Miller et al. (2021) determined that the widely used Perceived Stress Scale (PSS) measured stable characteristics of stress and was, therefore, more suitable for assessing long-term effects of interventions rather than short-term effects.

Although, both G-theory and Item Response Theory (IRT)/Rasch models are transcending limitations of CTT in investigating psychometric properties of ordinal scales, they focus on different aspects of measurement. While IRT and Rasch models concerned with unique contributions of individual items to the overall construct (e.g., item difficulty) and functioning of items response categories (Medvedev et al., 2020), G-theory examines the overall reliability and generalisability of assessment scores over time and sample population (Miller et al., 2021). Thus, both methods complement each other in establishing reliable and valid assessment instruments.

This study aimed to investigate stable and dynamic aspects of distress captured by the DASS-21 and to establish the reliability and generalisability of its assessment scores under pre-pandemic and pandemic conditions. Using data collected from the United States (pre-pandemic) and New Zealand (pandemic), a generalisability study (G-study) was conducted to investigate the overall reliability of the DASS-21 and its subscales, and to produce generalisability coefficients and indices reflecting state and trait aspects of the measure. Following the G-study, a decision study (D-study) was

conducted to optimise reliability and evaluate dynamic and enduring patterns across symptoms reflected by the individual items of the DASS-21. It was hypothesised that challenges associated with the COVID-19 pandemic conditions, would increase variability observed in the DASS-21 scores during pandemic conditions compared to pre-pandemic and would, therefore, help to identify dynamic aspects of distress.

METHOD

Participants

Participants were New Zealand and US university students. A total of 115 responses were collected on three occasions from the US students at approximately one-month intervals. The US sample consisted mainly of people who identified as White Americans (89.7%) and female (70%) students between the age of 18 and 22 years ($M = 18.9$, SD = .87). Unlike the US sample, from which data were collected between February and April 2019, the data from the New Zealand sample was collected between March and June 2020, before, during and after New Zealand's COVID-19 pandemic lockdown (Table S1.). The NZ sample contained a total of 114 responses collected on three occasions with approximately 14-day intervals. The NZ sample consisted of people who identified as NZ European (61%), Māori (16%; indigenous people of New Zealand), Asian (16%), Pasifika (3%) and other ethnicities (4%). These participants were predominantly female (85%) students between the age of 18 and 68 years ($M = 26.32$, SD = 9.92).

Procedure

The study was approved by the Ethics Committee of the University of Missouri for the US sample and the University of Waikato for the NZ sample, both following the ethical standards as set forth in the 1964 Declaration of Helsinki and its later amendments. All students partook in the study online as part of their course grades and were informed about the study before providing consent to participate.

Measures

The DASS-21 (Lovibond & Lovibond, 1995) consists of 21 items that assess levels of psychological distress. The self-report measure asks participants to indicate how much each statement applied to them over the past week using a 4-point Likert scale, with responses ranging from 0 ("did not apply to me at all"), to 3 ("Applied to me very much, or most of the time"). The Depression, Anxiety and Stress subscales of the DASS-21 each contain seven items. The DASS-21's subscales are calculated by summing the relevant items' responses and are interpreted according to a scoring table to indicate the severity of symptoms (e.g., moderate severity of depression falls between a total subscale score of 14 and 20).

Data analyses

IBM SPSS version 27 was used for every occasion in both samples to compute descriptive statistics, ICC, test–retest and Cronbach's alpha coefficients. In this study, we used mean imputation to replace missing responses (<.5%) on individual items following procedures used in other similar studies (Paterson et al., 2018). Although, there are more advanced imputation methods available, such as multiple imputations (Enders, 2010), using mean imputation is justified in this case because it was used to replace missing responses on individual scale items only when the responses were available for most scale items that already have high correlations with each other (e.g., $\alpha = .77–.93$). Moreover, this imputation was applied at each time point separately while the overall missing data was merely .5% and is negligible from statistical perspective, given that the data were normally distributed.

Following the guidelines of Medvedev et al. (2017) and Cardinet et al. (2011), G-theory analyses were conducted in four stages (described below) using Edu-G 6.1-e software (Swiss Society for Research in Education Working Group, 2006). For both the G-study and D-study, a random effects design was implemented, where the interactions between person (P), item (I) and occasion (O) are expressed as $P \times I \times O$. Both person and occasion facets were set as infinite to allow generalisability, while the items facet was fixed because the same items were used across all observations. Facets were defined using the trait perspective, where the person (P) facet is the object of measurement, and items and occasions are instrumentation facets (Cardinet et al., 2011). Variance attributed to the model by persons represents a trait component of the scale scores; variance attributed to the model by person and occasion interactions represents state components of the scale scores where dynamic changes over time can be observed.

Analysis of variance estimates were used to calculate the sum of squares and mean squares as well as the variance components for each facet and their associated interactions. Table S2 presents further information on this process, including the formulae used to calculate the contribution of each facet to the score and the interactions between person, item and occasion (Shavelson et al., 1989).

Next, relative G-coefficients (G_r) and absolute G-coefficients (G_a) were calculated (Cardinet et al., 2011). G-coefficients reflect the overall reliability or the generalisability of the measure in question (G coefficient = true person variance/true person

variance + error variance). G_r accounts for the relative sources of variance related to person computed as follows (Shavelson et al., 1989):

$$G_r = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\delta^2}; \sigma_\delta^2 = \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{po}^2}{n_o} + \frac{\sigma_{pio}^2}{n_i n_o}$$

where n_o is the number of occasions and n_i is the number of items.

G_a is an absolute model of measurement based on the test scores that accounts for all sources of error that may indirectly influence the absolute measure $\left(\sigma^2\Delta = \frac{\sigma_0^2}{n_o} + \frac{\sigma_i^2}{n_i} + \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{po}^2}{n_o} + \frac{\sigma_{io}^2}{n_i n_o} + \frac{\sigma_{pio}^2}{n_i n_o}\right)$ (Cardinet et al., 2011):

$$G_a = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\Delta^2}$$

Lastly, a D-study was conducted in which individual items and subscales were examined and modified by iteratively removing items to optimise reliability of the DASS-21 and its subscales as well as distinguish between state and trait components. A state component index (SCI) and trait component index (TCI) were also computed. These component indexes are coefficients that represent the proportion of variance accounted for a state and trait component in a scale, respectively. These formulae were developed by Medvedev et al. (2017) and are include below:

$$\text{SCI} = \frac{\sigma_{po}^2}{\sigma_{po}^2 + \sigma_p^2}; \text{TCI} = \frac{\sigma_p^2}{\sigma_{po}^2 + \sigma_p^2}$$

We have also computed variance components for each item, together with SCI values. Items with higher SCI (e.g., ≥.70) are considered as measuring a state to a larger extent, while items with higher TCI (e.g., ≥.70) are predominantly measuring trait aspects.

RESULTS

Descriptive statistics of the DASS-21 and subscales, separated by occasions and samples are presented in Table S3. DASS-21 scores were distributed close to normal, with values for skewness ranging from .72 to 1.18 and kurtosis from −.21 to .90 (West et al., 1995). Post hoc tests of the DASS-21's full scale indicated that Occasion 2 full scale, in addition to the Anxiety and Stress subscales were significantly different in the NZ sample and compared to Occasion 1. Internal consistency for the full scale and subscales were acceptable, with Cronbach's alpha of the total DASS-21 scale over three occasions ranging from .79 to .93 in the NZ sample and .77 to .93 in the US sample. Test–retest coefficients for the total scale ranged from .65 to .77, and for subscales values ranged from .54 to .76 across both samples. ICCs ranged from .59 to .74 across all subscales and samples. Across three occasions in the US sample, the proportion of participants who scored above the cut-off scores for severe disturbances ranged from 17 to 22% for depression, 21 to 26% for anxiety and 27 to 36% for stress, whereas the proportion of participants who met the cut-off score for extremely severe were 18 to 27%, 52 to 64% and 17 to 32%, respectively. Contrastingly, the NZ sample presented with a lower number of participants in severe category with proportions ranging from 7 to 11% for depression, 7 to 9% for anxiety and 10 to 11% for stress. Extremely severe proportions in the NZ sample ranged from 4 to 7% for depression, 7 to 11% for anxiety, and 4 to 0% for stress. Overall, the descriptive statistics from both samples indicated sufficient variability across all three occasions for the application of a G-theory analysis.

G-study

Table 1 includes variance components calculated in the G-study analyses. The results show that the DASS-21 total scale has excellent reliability and generalisability of scores across occasions and populations of both samples in measuring the overall psychological distress as a trait, reflected by over 90% of the true variance in the scores and G-coefficients ranging from .92 to .96. The overall error variance in the total scale scores was negligible and predominantly explained by person–occasion interaction representing individual state.

In the US sample, all three subscales of the DASS-21 demonstrated acceptable reliability and generalisability of scores with G_rs' ranging from .79 to .96. However, all three subscales appeared less reliable in the NZ sample during pandemic conditions as evidenced by G scores ranging from .61 to .71, which failed to meet the threshold for a reliable trait measure (≥.80; Arterberry et al., 2014). These results were in line with our expectation that variability of distress symptoms would be higher in pandemic conditions reflected by the overall lower G coefficients compared to the overall higher G-coefficients in the pre-pandemic sample. In the US sample, all three subscales of the DASS-21 demonstrated acceptable reliability and generalisability of scores with G_rs' ranging from .79 to .96. However, all three subscales appeared less reliable in the NZ sample during pandemic conditions as evidenced by G scores ranging from .61 to .71, which failed to meet the threshold for a reliable trait measure (≥.80; Arterberry et al., 2014). These results were in line with our expectation that variability of distress symptoms would be higher in pandemic conditions reflected by the overall lower G coefficients compared to the overall higher G-coefficients in the pre-pandemic sample.

The largest portion of measurement error in all subscales for both samples were the interactions between person and item ($P \times I$) and person, item and occasion

TABLE 1
Variance components and reliability estimates of the DASS-21's full scale and subscales across two samples

	DASS-21 total				*Depression*				*Anxiety*				*Stress*			
	US		*NZ*		*US*		*NZ*		*US*		*NZ*		*US*		*NZ*	
Facets	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^2	%
P	.162	.95	.082	92.00	.149	76.00	.067	57.00	.131	79.00	.091	69.00	.191	88.00	.062	60.00
I	.000	.00	.000	.00	.001	.50	.000	.00	.000	.00	.000	.00	.001	.56	.001	.56
O	.002	1.06	.002	2.00	.006	2.78	.008	6.92	.000	.13	.000	.37	.000	.00	.000	.00
PI	.000	.00	.000	.00	.019	9.50	.022	19.18	.014	8.51	.017	12.99	.008	3.72	.015	14.40
PO	.006	3.95	.006	6.00	.008	3.96	.005	3.96	.004	2.42	.005	3.50	.000	.06	.010	9.64
IO	.000	.00	.000	.00	.001	.62	.002	2.06	.002	1.37	.002	1.89	.002	.94	.002	1.64
PIO	.000	.00	.000	.00	.013	6.65	.013	10.92	.014	8.59	.016	12.28	.015	6.73	.014	13.72
G_r	.96 [.87, 1.05]		.94 [.85, 1.03]		.79 [.55, 1.03]		.63 [.39, .87]		.80 [.58, 1.02]		.71 [.48, .94]		.89 [.71, 1.07]		.61 [.37, .85]	
G_a	.95 [.85, 1.05]		.92 [.82, 1.02]		.76 [.50, 1.02]		.57 [.29, .85]		.79 [.57, 1.01]		.69 [.45, .93]		.88 [.68, 1.08]		.60 [.36, .84]	
SCI	.04		.07		.05		.07		.03		.05		.00		.14	
TCI	.96		.93		.95		.93		.97		.95		1.00		.86	

Note: *P* = person; *I* = Item; *O* = occasion; PI = person−item; PO = person−occasion; IO = item-person; PIO = person−occasion; G_r = G coefficient (relative); G_a = G coefficient (absolute); SCI = state component index; TCI = trait component index.

($P \times I \times O$) together explaining between 7 and 24% of the total variance. Additionally, when compared to the US sample, the NZ sample had more error variance from the person−item interaction, whereas the US sample had a larger portion of error variance from the person−item−occasion interaction. An interaction between person, item and occasion suggests that participants' responses to subscale items were influenced by assessment occasion (e.g., assessment environment). SCI coefficients for all scales across both samples ranged from .00 to .14 suggesting that variance attributed to the state component is relatively negligible compared to the trait variance.

D-study

Table S4 presents results from the D-study's individual item analysis, which includes variance components of person, person−occasion interaction, SCI and G_r. The most state-sensitive item in the scale was item 8 ("I felt that I was using a lot of nervous energy") from the stress subscale, with an SCI of .63 (US) and .70 (NZ), and G_r of .37 (US) and .30 (NZ), indicating that this item meets the threshold to be considered a state item (SCI ≥ .60; Cardinet et al., 2011; Medvedev et al., 2017). Contrastingly, most other items in both samples reflected predominantly enduring distress patterns. For instance, items 3, 4, 5, 6, 7, 10, 11, 18, 19 and 20 reflected the most stable distress symptoms across both samples with a $G_r \geq .60$. The remaining items had SCI below .50 in both samples, indicating that these items predominantly reflect enduring symptoms. Attempts to enhance the reliability of subscales by iteratively removing the most dynamic items for individual subscales and the total scale are included in Table S4. These analyses achieved no noticeable improvement of G-coefficients across both samples, suggesting that the DASS-21 and its subscales have optimal reliability in the current measurement design. There was no or little change when each of the three occasions was excluded from the generalisability analysis, indicating that the results were not affected by any specific occasion.

DISCUSSION

Using G-theory, this study investigated enduring and dynamic aspects of distress including depression, anxiety and stress symptoms operationalised by the DASS-21 under pre-pandemic and pandemic conditions. We used longitudinal data from two independent samples of approximately equal size, one collected in the US before COVID-19 and the other collected in NZ during the COVID-19 pandemic period. Across both samples, our results demonstrated that the DASS-21 full scale was reliable in measuring enduring symptoms of psychological distress, which is reflected by the excellent temporal reliability and generalisability of the total the DASS-21 scores across occasions and both samples' populations (G = .94–.96). This means that the DASS-21's total score accurately measures true differences between people with merely 4 to 6% of variance attributed to the measurement error, suggesting that the overall psychological distress has characteristic of a trait, which is relatively stable over time. Overall, the total distress scores were remarkably consistent across individuals over time, of which is in line with the results from a recent validation study conducted by Medvedev et al. (2020) where Rasch analysis demonstrated that the combined score of the DASS-21's subscales had robust reliability and internal validity. Attempts to enhance the reliability of the DASS-21 subscales by modifying item content did not yield any improvement, indirectly supporting the reliability of our primary results. Moreover, little change was observed

when one of the three occasions was excluded from the generalisability analysis, indicating that these results were not greatly influenced by any specific occasion, and the current measurement design was appropriate.

Although all three DASS-21 subscales showed good reliability and measured relatively enduring characteristics of depression, anxiety and stress, the DASS-21's subscales were found to be less reliable and measured more dynamic characteristics of depression, anxiety and stress under pandemic conditions (NZ sample) compared to pre-pandemic conditions (US sample). Most notably the stress subscale in the NZ sample had G coefficients below acceptable levels to be considered as a reliable measure of either state or trait (Arterberry et al., 2014). These findings support our hypothesis that the variability of depression, anxiety and stress symptoms would increase during pandemic conditions. It should be noted that our samples are not directly comparable as they were drawn from different populations and under different conditions and our study has a quasi-experimental nature that allowed us to examine the assessment of distress under natural conditions associated with both pandemic and everyday life.

The results showed that the individual subscales were mostly affected by measurement error due to person–item interaction. Error variance for person–item interaction ranged from 3.72 to 19.18% across all subscales, with the overall higher values observed in the NZ sample during the pandemic. Measurement error associated with person–item interaction suggests differential item functioning and/or interpretation, meaning that interpretation of items varied from person to person resulting in less consistent answers. Furthermore, larger proportions of error variance associated with person–item interaction were observed for depression, anxiety and stress subscales in the NZ sample while the US sample's participants answered questions more consistently. This indicated that the NZ participants may be understanding and answering items in a less consistent way compared to the US participants. However, person–item and person–item–occasion interaction errors were only observed at the individual subscale level and did not impact on the overall DASS 21 total score. Additionally, a large portion of measurement error was due to the interaction between person, item and occasion. Error variance for person–item–occasion interaction ranged from 6.65 to 13.72%, suggesting that subscales' scores were influenced by both the occasion that the questionnaire was administered and the participant's personality, which in turn affected how the questions were interpreted.

It should be noted that differential interpretation and item functioning can be overcome by transforming the ordinal scores into interval level data using the Rasch conversion algorithm, which was already developed for the DASS-21 with the New Zealand population (Medvedev et al., 2020). Future studies should use Rasch or IRT methodology to establish scale invariance between the US and NZ populations. However, the overall higher amount of error due to person–item interaction in the NZ sample compared to the US sample led us to tentatively speculate that the increased variances seen in person–item and person–item–occasion interaction were likely due to the fact that NZ data was collected during the COVID-19 pandemic and lockdowns, which might affect how people perceived their distress symptoms. For instance, a recent study conducted by Hartstone and Medvedev (2021) found a reduction in anxiety and stress levels during New Zealand's COVID-19 lockdown. It is possible that the larger person–item interaction seen in the NZ sample is associated somewhat with the overall lower anxiety and stress as reported by Hartstone and Medvedev (2021). On a different note, the US sample had a significantly larger proportion of participants who met the over-cut-off scores for severe and extremely severe disturbance across all three subscales, which suggest more familiarity with distress symptoms and hence more consistency in responding. In addition, such observations could be explained by differences in lifestyle between countries.

A key finding in our study was the remarkable stability of the overall scale scores across both conditions and samples, indicating that participants' overall distress levels were likely to remain relatively stable over time. In fact, the error variance associated with both person–item and person–item-occasion interaction naturally decreases as the number of scale items increases and similar trends were demonstrated across different studies and samples (Medvedev et al., 2017; Miller et al., 2021; Truong et al., 2020). An example of this can be seen in Table 1 where the US and NZ total scale measurement error for person–item–interaction and person–item–occasion become negligible when items were pooled together, while the overall G_a and G_r increased indicating enhancement of reliability and generalisability of assessment scores. Therefore, the high stability of the total scale remains an interesting finding considering that the depression, anxiety and stress operationalised by the DASS-21 were subject to change over time to a larger extent, which was further emphasised in the NZ sample. Moreover, given that symptoms of depression, anxiety and stress share such a close relationship and often co-occur or co-exist together, it is likely that the observed variability represents shifts between the three constructs and is reflective of their strong relationship (Zhang et al., 2020).

Studies using G-theory to investigate psychometric scales with distress-related constructs have found similar, yet different findings compared to our study. For instance, Miller et al. (2021) study on the PSS-10 reported predominantly enduring symptoms of perceived stress with stable G_r's of .86 across control and intervention groups, whereas the DASS-21's Stress scale had significantly larger differences and variation of reliability between samples, with a G_r of .89 for the US sample and .61 for the NZ sample. An additional but notable

similarity is that both the DASS-21's Stress subscale and the PSS-10 were found to measure stable aspects of stress. Another example is Paterson et al.'s (2018) examination of the CDI-10 in which they demonstrated that the CDI-10 measured both stable and dynamic aspects of childhood depression with a G_r value of .79 and TCI of .97, which was similar to the DASS-21 Depression subscale, which had G_r values ranging from .63 to .79 and TCIs from .95 to .93. Forrest et al. (2021) applied G-theory to the widely used State and Trait Anxiety Inventory (STAI; Spielberger, 1983) and demonstrated using two independent samples that both subscales of the STAI measure trait anxiety with G-coefficients ranging from .84 to .96, which is consistent with our results in the non-pandemic condition.

Limitations

A notable limitation of our study was a degree of homogeneity in terms of participants' age and participation in higher education. Additionally, data was collected from different countries (US and NZ) and under different conditions (pre-pandemic and pandemic) which did not allow for direct valid comparison between samples, and we can only speculate as to why the observed increases in variability seen in the NZ sample were due to the differences between samples or conditions. Moreover, it would have been more optimal to collect data from the same location pre-pandemic and during pandemic conditions as some differences observed in the DASS-21 may be also attributed to the location and time of data collection. Nevertheless, this study aimed to test the psychometric properties of the DASS-21 and to determine how the scale was likely to behave under different conditions and sample populations. For this study, the diversity of sample populations and conditions were desirable and served to evaluate the reliability and generalisability of DASS-21 scores as well as to distinguish enduring and dynamic aspects of psychological distress.

Conclusion

In summary, the findings from this study indicate that the overall symptoms of psychological distress operationalised by the DASS-21 are relatively stable during the pandemic and normal conditions, and the total DASS-21 assessment scores can be generalised across sample populations and occasions. However, the depression, anxiety and stress symptoms captured by the DASS-21 subscales were notably less stable suggesting shifts across depression, anxiety and stress levels in both samples further amplified during pandemic conditions in NZ. High measurement error associated with person−item−occasion and person−item interaction found for individual subscales suggests differential item functioning, which can be mitigated by utilising the DASS-21's total scale not affected by such errors. These findings may have important implications for the reliability of assessment, diagnostic and treatment of affective symptoms in normal and pandemic conditions associated with COVID-19. Overall, these findings suggest that differential item functioning is more likely if using short scales containing less than 10 items and longer scales would be preferable where high precision of measurement is required or measurement is likely affected by external circumstances such as a pandemic.

Data Availability Statement

This study is preregistered at the Center for Open Science OSF accessible through the follow link: osf.io/nq7zu. The data that support the findings of this study are also available on request from the corresponding author.

Manuscript received October 2021
Revised manuscript accepted March 2023

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Alert levels by date in New Zealand during the duration of the current study.

Table S2. Formulas used to estimate the effects for all facets presented by observed scores X and related variance components (Shavelson et al., 1989).

Table S3. Means, standard deviation (SD), Cronbach's alpha, test−retest coefficients, intraclass correlation (ICC), 95% confidence intervals (CI), with proportions of participants who meet cut-off values for severe and extremely severe for the DASS-21 over three occasions across two samples.

Table S4. Decision study indicating state component index for all items in the DASS-21 across both samples with full scale and subscale modifications.

REFERENCES

Alzueta, E., Perrin, P., Baker, F., Caffarra, S., Ramos-Usuga, D., Yuksel, D., & Arango-Lasprilla, J. (2021). How the COVID-19 pandemic has changed our lives: A study of psychological correlates across 59 countries. *Journal of Clinical Psychology*, *77*(3), 556−570. https://doi.org/10.1002/jclp.23082

Arterberry, B. J., Martens, M. P., Cadigan, J. M., & Rohrer, D. (2014). Application of generalizability theory to the big five inventory. *Personality and Individual Differences*, *69*, 98−103. https://doi.org/10.1016/j.paid.2014.05.015

Bloch, R., & Norman, G. (2012). Generalizability theory for the perplexed: A practical introduction and guide: AMEE guide no. 68. *Medical Teacher*, *34*(11), 960−992. https://doi.org/10.3109/0142159X.2012.703791

Cardinet, J., Pini, G., & Johnson, S. (2011). *Applying generalizability theory using EduG*. Routledge Academic.

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: Aliberation of reliability theory. *The British Journal of Statistical Psychology*, *16*(2), 137–163. https://doi.org/10.1111/j.2044-8317.1963.tb00206.x

Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.

Forrest, S. J., Siegert, R. J., Krägeloh, C. U., Landon, J., & Medvedev, O. N. (2021). Generalizability theory distinguishes between state and trait anxiety. *Psychological Assessment*, *33*, 1080–1088. https://doi.org/10.1037/pas0001060

Hartstone, J. M., & Medvedev, O. N. (2021). The role of mindfulness and life satisfaction in psychological distress during the COVID-19 lockdown in New Zealand: A quasi-experimental study. *Mindfulness*, *12*(11), 2693–2706. https://doi.org/10.1007/s12671-021-01731-4

Kircanski, K., LeMoult, J., Ordaz, S., & Gotlib, I. H. (2016). Investigating the nature of co-occurring depression and anxiety: Comparing diagnostic and dimensional research approaches. *Journal of Affective Disorders*, *216*, 123–135. https://doi.org/10.1016/j.jad.2016.08.006

Kwong, A. S. F., Pearson, R. M., Adams, M. J., Northstone, K., Tilling, K., Smith, D., Fawns-Ritchie, C., Bould, H., Warne, N., Zammit, S., Gunnell, D. J., Moran, P., Micali, N., Reichenberg, A., Hickman, M., Rai, D., Haworth, S., Campbell, A., Altschul, D., … Timpson, N. J. (2021). Mental health during the COVID-19 pandemic in two longitudinal UK population cohorts. *The British Journal of Psychiatry*, *218*(6), 334–343. https://doi.org/10.1192/bjp.2020.242

Lee, D. (2019). The convergent, discriminant, and nomological validity of the depression anxiety stress scales-21 (DASS-21). *Journal of Affective Disorders*, *259*, 136–142. https://doi.org/10.1016/j.jad.2019.06.036

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales* (2nd ed.). Psychology Foundation of Australia.

Medvedev, O. N., Krägeloh, C. U., Narayanan, A., & Siegert, R. J. (2017). Measuring mindfulness: Applying generalizability theory to distinguish between state and trait. *Mindfulness*, *8*(4), 1036–1046. https://doi.org/10.1007/s12671-017-0679-0

Medvedev, O. N., Krägeloh, C. U., Titkova, E. A., & Siegert, R. J. (2020). Rasch analysis and ordinal-to-interval conversion tables for the depression, anxiety and stress scale. *Journal of Health Psychology*, *25*(10–11), 1374–1383. https://doi.org/10.1177/1359105318755261

Medvedev, O. N., Norden, P. A., Krägeloh, C. U., & Siegert, R. J. (2018). Investigating unique contributions of dispositional mindfulness facets to depression, anxiety, and stress in general and student populations. *Mindfulness*, *9*(6), 1757–1767. https://doi.org/10.1007/s12671-018-0917-0

Miller, Y. R., Medvedev, O. N., Hwang, Y.-S., & Singh, N. N. (2021). Applying generalizability theory to the perceived stress scale to evaluate stable and dynamic aspects of educators' stress. *International Journal of Stress Management*, *28*(2), 147–153. https://doi.org/10.1037/str0000207

Osman, A., Wong, J. L., Bagge, C. L., Freedenthal, S., Gutierrez, P. M., & Lozano, G. (2012). The depression anxiety stress scales-21 (DASS-21): Further examination of dimensions, scale reliability, and correlates. *Journal of Clinical Psychology*, *68*(12), 1322–1338. https://doi.org/10.1002/jclp.21908

Paterson, J., Medvedev, O., Sumich, A., Tautolo, E., Krägeloh, C., Sisk, R., McNamara, R., Berk, M., Narayanan, A., & Siegert, R. (2018). Distinguishing transient versus stable aspects of depression in New Zealand Pacific Island children using generalizability theory. *Journal of Affective Disorders*, *227*, 698–704. https://doi.org/10.1016/j.jad.2017.11.075

Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist*, *44*, 599–612.

Spielberger, C. D. (1983). *Manual for the state-trait anxiety inventory*. Consulting Psychologists Press.

Swiss Society for Research in Education Working Group. (2006). *EDUG user guide*. IRDP.

Truong, Q. C., Krägeloh, C. U., Siegert, R. J., Landon, J., & Medvedev, O. N. (2020). Applying generalizability theory to differentiate between trait and state in the five facet mindfulness questionnaire (FFMQ). *Mindfulness*, *1–11*, 953–963. https://doi.org/10.1007/s12671-020-01324-7

West, S. G., Finch, J. F., & Curran, P. J. (1995). *Structural equation modeling: Concepts, issues and applications SAGE publications*. Thousand Oaks.

World Health Organization. (2020). *COVID-19 disrupting mental health services in most countries, WHO survey*. https://www.who.int/news/item/05-10-2020-covid-19-disrupting-mental-health-services-in-most-countries-who-survey

Zhang, J., Huen, J. M. Y., Lew, B., Chistopolskaya, K., Talib, M. A., Siau, C. S., & Leung, A. N. M. (2020). Depression, anxiety, and stress as a function of psychological strains: Towards an etiological theory of mood disorders and psychopathologies. *Journal of Affective Disorders*, *271*, 279–285. https://doi.org/10.1016/j.jad.2020.03.076