



Research Paper

Application of generalizability theory to evaluate the Kessler Psychological Distress Scale and distinguish between enduring and dynamic distress

Vivienne Yu-X Yan^a, Katya Numbers^b, Perminder S. Sachdev^b, Henry Brodaty^b, Oleg N. Medvedev^{a,*}

^a University of Waikato, School of Psychology, Hamilton, New Zealand

^b Centre for Healthy Brain Ageing (CHeBA), Discipline of Psychiatry and Mental Health, University of New South Wales, Sydney, Australia

ARTICLE INFO

Keywords:

Distress
Older adults
K-10
G-theory
Enduring and dynamic symptoms
Psychometric measurement

ABSTRACT

Objectives: Nowadays, the number of older people is increasing rapidly, both in absolute figures and as a proportion of the population, which makes the maintenance of psychological well-being among aging population imperative. Neuropsychological distress may promote cognitive impairment and impact on the health of older people, which makes accurate assessment of distress an important clinical and research issue. The Kessler Psychological Distress Scale (K-10) is a widely used instrument used to measure individual distress. However, the ability of K-10 to distinguish between enduring and dynamic distress symptoms, and the generalizability of its scores, have not been investigated in older populations using appropriate methodology.

Method: Generalizability theory (G-theory) was applied to differentiate enduring and dynamic distress and examine the reliability of K-10 in a sample of 201 adults (43% males) aged 70 to 90 years old who participated in Sydney Memory and Ageing Study. The data were collected biennially over ten years.

Results: The K-10 scale showed strong reliability ($G_a=0.81$, $G_r=0.89$) in assessing enduring distress and its assessment scores were generalizable across occasions and older adults. Most of distress symptoms represented by K-10 items were enduring.

Limitations: Generalizability of these findings may be limited to older adults.

Conclusions: The K-10 appears more suitable to evaluate enduring distress in aging populations over time, which is important for interventions targeting older adults' mental health given its scores will likely capture enduring changes in neuropsychological health before and after interventions.

Psychological distress, characterized by intense emotional suffering often associated with symptoms of depression and anxiety, is notably present in older individuals, significantly affecting their cognitive functions and overall well-being (Drapeau et al., 2012; Ridner et al., 2004; Weissman et al., 2015; Paül and Santos, 2009; Doris et al., 2004; Voyer et al., 2005). This distress can be transient or enduring, with the latter being associated with severe health conditions and impaired life satisfaction, emphasizing the necessity for precise differentiation and timely interventions (Forrest et al., 2021; Miller et al., 2020; Medvedev et al., 2017).

The rising global older adult population underscores an urgent need for integrated health services focused on promoting psychological health and well-being during aging (World Health Organization, 2022; Australian Bureau of Statistics, 2022). However, the effective assessment and treatment of psychological distress in older adults is often

hampered by underestimation of symptoms, partly due to societal stigmatization and clinical misconceptions regarding the natural aging process (Conner et al., 2010; Nurse Key, 2022; Moulton et al., 2020).

To address this, the Kessler Psychological Distress Scale (K-10), a widely recognized screening tool validated across diverse populations, has played a critical role in identifying and evaluating levels of psychological distress (Kessler et al., 2002; 2003; Brook et al., 2006; Easton et al., 2017). This scale, with high accuracy in identifying affective disorders and sound psychometric properties, offers a promising avenue for detailed psychological assessment (Smout, 2019; Brooks et al., 2006).

The K-10 has demonstrated remarkable validity, reflected in its efficacy in predicting symptoms of depression and anxiety across various studies, showcasing its strong criterion and convergent validity (Lu et al., 2014; McShane et al., 2016; Lace et al., 2019). Its reliability is

* Corresponding author at: School of Psychology, University of Waikato Private Bag 3105, Hamilton 3240, New Zealand.

E-mail address: oleg.medvedev@waikato.ac.nz (O.N. Medvedev).

<https://doi.org/10.1016/j.jadr.2023.100668>

Received 19 May 2023; Received in revised form 13 September 2023; Accepted 8 October 2023

Available online 10 October 2023

2666-9153/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

underscored by substantial internal consistency across different versions and acceptable temporal stability (Kessler et al., 2003; Arnaud et al., 2010; Easton et al., 2017; Thelin et al., 2017; Merson et al., 2021). However, current reliability assessments might overlook potential measurement errors associated with individual items and their interactions over time, pointing to limitations of existing research and highlighting the necessity for a more nuanced approach to assessing the K-10's reliability (Medvedev and Siegert 2022; Welsh et al., 2020).

G-theory, developed by Cronbach et al. (1963), offers a comprehensive approach to understanding reliability by considering various sources of variability potentially affecting assessment scores. This theory goes a step further than classical test theory by accounting for the different sources of variance and their interactions, facilitating a deeper understanding of the reliability of psychometric measurements (Vispoel et al., 2018; Cronbach et al., 1963).

Applying G-theory to the assessment of K-10 promises a detailed insight into the enduring and dynamic aspects of psychological distress, delineating between state and trait characteristics captured by the tool (Arterberry et al., 2014; Paterson et al., 2018; Medvedev and Siegert 2022). This analysis aids in understanding the tool's capacity to differentiate between transient and persistent symptoms of distress, which is essential for accurate diagnosis and effective intervention planning (Medvedev et al., 2017).

This study leverages G-theory to scrutinize the generalizability of K-10 scores across different occasions, aiming to discern the enduring and dynamic aspects of psychological distress in older adults using data from the Sydney Memory and Ageing Study (MAS). The analysis follows a two-step approach involving a G-study and a D-study, promising a rigorous assessment of the K-10's efficacy while offering insights into potential scale modifications to enhance the reliability and precision of this essential psychometric tool (Shavelson et al., 1989; Medvedev et al., 2021; Truong et al., 2020).

1. Method

1.1. Participants

In 2005, a comprehensive selection process was initiated to recruit individuals aged between 70 and 90 years for the Sydney Memory and Ageing Study (MAS; Sachdev et al., 2010). Initially, a substantial pool of 8914 community-dwelling seniors residing in the Eastern Suburbs of Sydney, Australia, was identified using the electoral register. This strategic approach aimed to involve a diverse participant group and foster a detailed exploration of memory and aging dynamics in the Sydney populace. To maintain a stringent focus on the study's objectives, the research team imposed several prerequisites for participation. Prospective participants were required to have a proficient understanding of English to facilitate seamless communication during psychometric assessments and while responding to self-report questionnaires. Additionally, individuals willing to provide informed consent were considered for further evaluation.

The exclusion parameters were delineated to omit individuals with a previously documented dementia diagnosis, those suffering from significant psychological or neurological disorders, or individuals battling progressive malignancy. A critical facet of the selection process involved assessing participants through the Mini-Mental State Examination (MMSE; Folstein et al., 1975). A predetermined cutoff score below 24 led to the exclusion of several participants to ensure a homogeneous group in terms of cognitive functioning. Moreover, individuals identified with dementia at the onset of the study were not included in the research cohort. Following a meticulous selection process governed by the criteria outlined above, a cohort of 1037 individuals emerged as eligible participants from the initial pool of 8914 potential candidates. This robust methodology ensured the assembly of a participant group bearing relevance and coherence with the study's focal points, primed for analytical pursuits exploring variations in K10 scores over time and

across different items. For an enriched understanding of the recruitment strategy and an overview of the baseline demographics, readers are encouraged to refer to the detailed exposition presented in a previously published work (Sachdev et al., 2010).

For the present study, participants were further excluded if they did not have data at all five Waves ($n = 201$) of the study (i.e., baseline to 10-year follow-up). The power analysis was conducted using G*power software. The current sample ($n = 201$) was larger than the minimum required sample size ($n = 187$) for repeated ANOVA over five times necessary to achieve the power ($1-\beta$) of 0.95 to detect very small effect size of 0.10 using conventional p -value of 0.05. Baseline demographics characteristics for the full sample ($n = 1037$) and current sample ($n = 201$) are presented in Table 1. It can be seen that demographic characteristics of the full sample and sub-sample are comparable to a high degree including age, distress, proportion of males and females and education levels. Although, there was a significant age difference between the full sample and the study sample, the mean difference was merely 1.8 years and can be considered as negligible. The CONSORT diagram in Fig.1 illustrates the number of participants who provided K-10 data at each of the five occasions including those diagnosed with dementia. The K-10 scale was not administered at Wave 5 (8-year follow-up), though it was available at Wave 6 (10-year follow-up); thus, participants for the current sample were required to have K-10 data at Waves 1, 2, 3, 4, and 6. All participants provided informed consent, and ethical approval for the study was obtained from the University of New South Wales Human Ethics Review Committee (HC 05,037, 089,382, 14,327).

2. Measures

The K-10 (Kessler et al., 2003) is a well validated scale used to measure psychological distress. The scale includes ten questions describing key symptoms of distress (e.g., did you feel so nervous that nothing could calm you down?). In the current sample, the K-10 demonstrated strong internal consistency with both Cronbach's Alpha and McDonald's Omega ranging from 0.81- 0.84 across all five assessment occasions. For the K-10, participants respond to each question using five response categories ranging from 1 to 5 (None of the time=1, A little of the time=2, Some of the time=3, Most of the time=4, All the time=5). The total score is calculated by summing individual item scores. K-10 scores can be interpreted as follows (Kessler et al., 2003): 10–19 unlikely unwell, 20–24 prospectively to have a mild distress, 25–29 prospectively to have a moderate distress and 30–50 prospectively to have a severe distress.

Table 1

Sample demographics including full sample and G-study sub-sample.

Demographic characteristics	Original sample $n = 1037$	G-study sample $n = 201$	p -value
Age: Mean (SD)	78.84 (4.84)	77.02 (4.39)	<.01 ^a
Sex			
Male	465(44.8%)	86(43.0%)	.59 ^b
Female	572(55.2%)	115(57.0%)	
Education			.82 ^b
Primary School	26	5	
Incomplete High School	411	73	
Completed High School	142	27	
Incomplete Tertiary	35	11	
Completed Tertiary	311	65	
Incomplete High School + Certificate/Diploma	57	10	
Complete High School + Certificate/Diploma	55	10	
Distress K-10 scores: Mean (SD)	13.40 (3.62)	12.84 (2.99)	.05 ^a

Note:

^a t -test.

^b χ^2 -test.

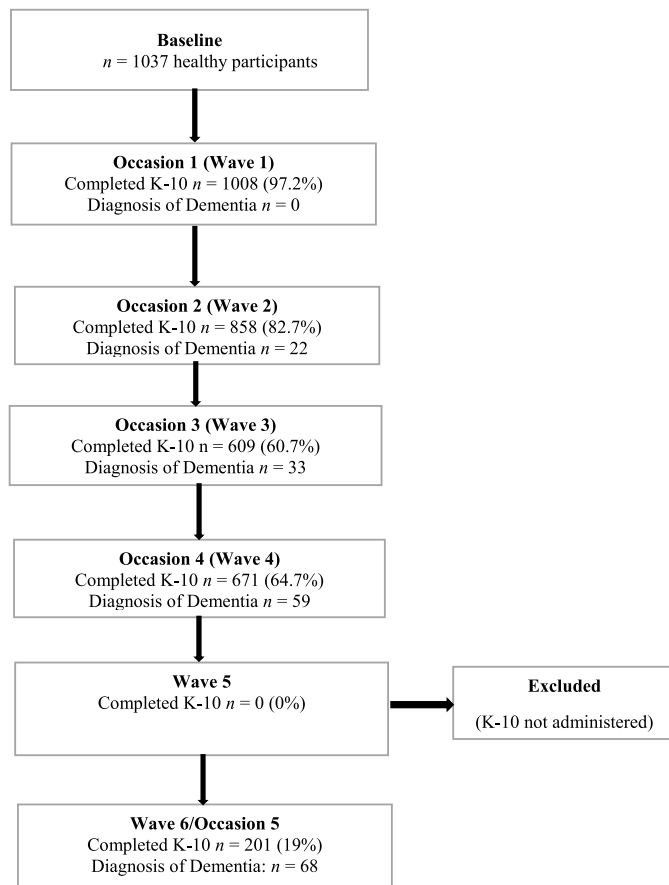


Fig. 1. CONSORT diagram including the number of participants who completed the K-10 at five Waves/Occasions respectively, and the number of diagnosed with dementia at each occasion.

3. Data analyses

In addressing missing data in the K10 scores, which was below 0.01%, we utilized a person mean substitution method, separately applying it for each wave to prevent any alteration in the temporal patterns evident in the data. This method was chosen due to the negligible amount of missing data, pertaining to between 1 and 5 individuals across different waves, with none having more than 5 missing scores out of 10. Importantly, supplementary analyses employing different strategies for handling missing data affirmed the robustness of our results, demonstrating no significant differences compared to the findings achieved through mean substitution. This strategy, thereby, assures the reliability and validity of the outcomes presented, providing a firm grounding for the analytical narrative drawn from the dataset (Huisman, 2000).

Descriptive data was analyzed using the IBM SPSS 28 software and involved repeated measures ANOVA to compute and analyze the overall distress scores over time. The reliability of the scale scores was computed including internal consistency and test-retest. G-theory analyses were conducted by applying EduG6.1-e software (Swiss Society for Research in Education Working Group, 2006). Both the G-study and D-study investigated true variance of Person (P) as an object of assessment with two crossed facets: by I (Item) and by O (occasion); this equation can be represented as P x I x O (Cardinet et al., 2011). Therefore, the generalizability of K-10 scores was examined across the sample population (P) and occasions (O), which were defined as infinite facets, and across I (Item) facets, which were fixed (Forrest et al., al., 2021).

Fig. 2 is an illustration of the variance components: person (P), item

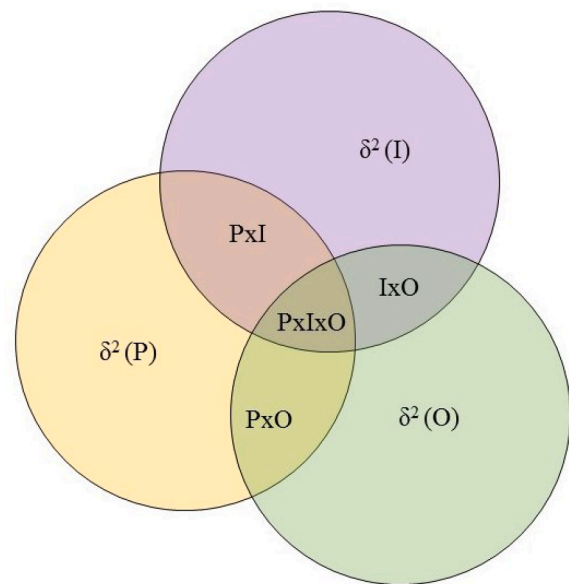


Fig. 2. Illustration of variance (δ^2) sources and their interactions for person by item by occasion measurement (P x I x O) design.

(I), occasion (O), and their interactions. Dynamic distress (i.e., their state) depends on both a person’s predisposition to experience distress (i.e., their trait) and current circumstances that may trigger distress, and is commonly defined as variance due to an interaction between person and occasion (P x O). The interaction between person and item (P x I) indicates potential error variance due to different understandings of the same questions by different people. The interaction between item and occasion (I x O) represents the error variance of the assessment occasions on how people understand scale items. The interaction among person, item, and occasion (P x I x O) represents the overall error associated with how a person responds to individual items and the influence of the occasion on that response. This complex interaction (P x I x O) is represented by the Ven diagram below (Fig. 2).

G-theory was used to compute the relative G-coefficient (Gr), and absolute G-coefficient (Ga), which reflect the overall reliability and generalizability of the K-10 assessment scores in this sample. The relative-coefficient only accounts for the variance related to the person (Shavelson et al., 1989), such as the interaction between person and occasions or person and item. The formulation of Gr according to Shavelson et al. (1989) is below:

$$Gr = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\delta^2}$$

where σ_p^2 represents the relative variance related to person defined as $\sigma_p^2 = \sigma_{pi}^2 + \sigma_{po}^2 + \sigma_{pio}^2$; and error variance is defined as $\sigma_\delta^2 = \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{po}^2}{n_o} + \frac{\sigma_{pio}^2}{n_i n_o}$; where n_i is number of scale items, n_o = number of occasions. If the Gr coefficient is ≥ 0.80 , it describes a strong reliability of the scale in measuring traits (Chalmers et al., 2021). Ga is similar to the Phi (Φ) coefficient and Shavelson et al. (1989) formulated it as follows:

$$Ga \simeq \Phi = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\Delta^2}$$

Ga calculates both direct and indirect factors such as items and occasions ($\sigma^2\Delta = \frac{\sigma_o^2}{n_o} + \frac{\sigma_i^2}{n_i} + \frac{\sigma_{io}^2}{n_i} + \frac{\sigma_{po}^2}{n_o} + \frac{\sigma_{pi}^2}{n_i n_o} + \frac{\sigma_{pio}^2}{n_i n_o}$), which influence the absolute measurement (Cardinet et al., 2011). A good generalizability is explained with Ga coefficient ≥ 0.80 and values ≥ 0.70 are considered adequate (Chalmers et al., al.,2021). The state component index (SCI) and trait component index (TCI) were computed in the D-study to

explain the dynamic and enduring components as developed by (Oleg et al.,2017):

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

The SCI = 0.50 suggests that dynamic and enduring components have the same proportion; the SCI > 0.60 indicates a valid measure of dynamic characteristics (P x O), and TCI > 0.60 shows a good validity of trait measurement (Medvedev et al.,2017).

4. Results

The K-10 scores were distributed close to normal with the most participants unlikely to be unwell ($M = 12.84$; $SD=2.99$; Table 3) and remain well across 5 assessment occasions as illustrated in Fig. 3. There were no significant differences between mean stress scores across the first four occasions. However, there was a noticeable increase of stress between occasions four and five, which was statistically significant as indicated by repeated measures ANOVA ($F(200, 1) = 5.72, p = .001$ and subsequent post-hoc test ($p = .005$). This represents the overall occasion effect on the full sample as reflected by variance attributed to (O) occasion in G-theory. It should be noted that the statistically significant increase of distress at occasion 5 was below the cut-off point for mild distress.

4.1. G study

Table 2 includes variance components for P (person), O (occasion) and I (item) and their interactions as estimated by the G-Study. As shown, the P (person) variance explained 89.0% of the variances in the data. Therefore, the K-10 scale demonstrated a high reliability ($G_a = 0.81, G_r = 0.89$) in terms of the generalizability of assessment scores across the population and across occasions. The P (person) x O (occasion) interaction indicates an individual state or dynamic aspect of stress which comprised 10.5% of variance in that assessment scores and represents the largest source of error followed by O (occasion). The occasion explained 8.5% of the variance and reflects significant increase of the overall stress levels of the sample over time, which was only significant between occasions 4 and 5.

Considering the uncertainty surrounding G coefficients is also important, given the relatively modest sample size of the study. The $G_r=0.89$; 95% CI [0.75,1.03] and $G_a=0.81$; 95% CI [0.63,0.99] reflect the generalizability of the K-10 assessment scores with 95% confidence intervals, which indicate the potential range within which the true G-values may lie. This approach of emphasizing the confidence intervals aims to confer a robust understanding of the findings while delineating the associated uncertainty. It facilitates a deeper appreciation of the

potential fluctuations in the coefficients, offering a window into the variability that might be expected in a wider population, and hence should be considered in interpreting the results.

4.2. D-study

The D-study investigated different components of the K-10 (e.g. items) to optimize its reliability and to evaluate enduring and dynamic components in individual distress symptoms. Table 3 shows variance components for each individual item (representing a symptom) including P (person), O (occasion), P x O (person-occasion interaction), and TCI, which reflects a proportion of variance associated with enduring patterns of distress relative to dynamic patterns of distress as captured by the scale/item representing a symptom. The TCI scores range from 0.55 to 0.79 which represents a higher proportion of trait variance compared to state variance, meaning that all individual items measure enduring distress symptoms to the larger extent. At the same time, all G coefficients for individual items, with exception of item 5 “How often did you feel restless?”, were above 0.60, which consistently reflects characteristics of enduring symptoms. To explore whether modifications to the K-10 improved the reliability of the scale, we removed items with the lowest TCI (i.e., items 5 & 6), both individually and together, to determine whether the modified scale improved the overall reliability of the K-10 compared to the original full-scale version. Similarly, we examined whether the O (occasion) impacted on the results, by removing assessment occasions one at a time, though no noticeable changes in G coefficients were observed.

5. Discussion

In this study, we applied G-theory analyses to explore the reliability and generalizability of the K-10 scale, assessing the enduring and dynamic aspects of psychological distress in an aging population. We corroborate previous work underscoring the K-10 scale’s validity and reliability (Berle et al., 2010; Hobbs et al., 2018; Kessler et al., 2002; Merson et al., 2021), presenting new insights into the distress symptoms’ dynamics in older adults. Our analyses demonstrated high generalizability of the K-10 assessment scores across populations and assessment instances ($G_r=0.89; G_a=0.81$), asserting the scale efficacy in capturing stable traits of psychological distress. Notably, the person variance explained a substantial part of the total variance (81%), indicating the K-10’s ability to discern genuine individual differences in distress levels. While respecting the potential uncertainties indicated by the confidence intervals associated with the G coefficients, our findings remain substantial and contribute to the K-10 scale’s robust characterization. Drawing attention to this uncertainty not only adheres to a rigorous analytical practice but also fosters a balanced interpretation of the

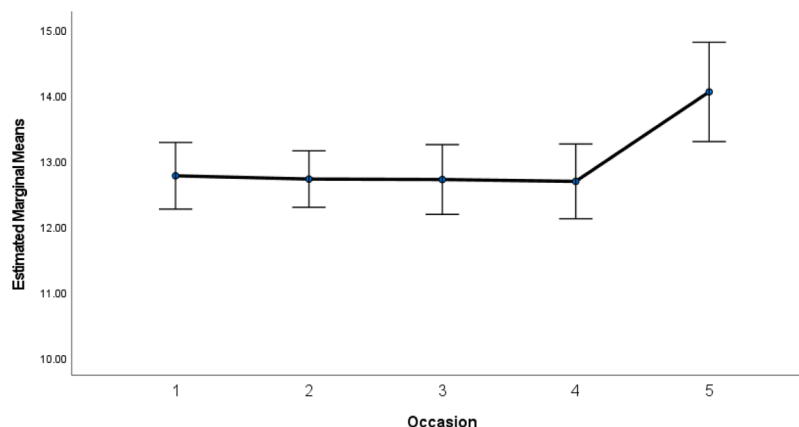


Fig. 3. Mean stress scores on five occasions including 95% confidence intervals.

Table 2

G-study results for the K-10 containing variance estimates including SE (standard errors), G-coefficients (Gr=relative and Ga=absolute) for P (Person) by I (Item) and by O (Occasion) observation design.

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.038		89.0	81.0
	I		(0.000)	0.0
	O		0.004	8.5
	PI	(0.000)	0.0	(0.000)	0.0
	PO	0.005	11.0	0.005	10.5
	IO		(0.000)	0.0
	PIO	(0.000)	0.0	(0.000)	0.0
Sum of variances	0.038		0.005	100%	0.009	100%
Standard deviation	0.195		Relative SE: 0.070		Absolute SE: 0.094	
Gr	0.89; 95% CI [0.75,1.03]					
Ga	0.81; 95% CI [0.63,0.99]					

Note: Grand mean =1.30; Absolute variance components in the last column on the right-hand side were computed out of 100%. Numbers in brackets indicate statistically negligible variance components.

Table 3

D-study results including variance components of P (Person), O (Occasion) and interaction between Persons and Occasion (P × O), TCI (Trait Component Index) and Generalizability Coefficient (Gr × Ga) for each item in the K10 scale.

No	Item content	P	O	P×O	TCI	Gr/a
1	How often did you feel tired out for no good reason?	0.089	0.016	0.054	0.62	0.62/ 0.56
2	How often did you feel nervous?	0.062	0.012	0.039	0.61	0.61/ 0.55
3	How often did you feel so nerves that nothing calm you down?	0.086	0.021	0.054	0.61	0.61/ 0.53
4	How often did you feel hopeless?	0.080	0.010	0.037	0.69	0.69/ 0.63
5	How often did you feel restless?	0.071	0.021	0.058	0.55	0.55/ 0.47
6	How often did you feel so restless that nothing calms you down?	0.055	0.010	0.036	0.60	0.60/ 0.54
7	How often did you feel depressed?	0.133	0.017	0.050	0.73	0.73/ 0.66
8	How often did you feel that everything was an effort?	0.128	0.007	0.033	0.79	0.79/ 0.76
9	How often did you feel so sad that nothing could cheer you up?	0.112	0.021	0.045	0.71	0.71/ 0.63
10	How often did you feel worthless?	0.107	0.007	0.036	0.75	0.75/ 0.71
	Scale modifications					
	Removing Item 5	0.038	0.003	0.005	0.88	0.86/ 0.80
	Removing Item 6	0.040	0.004	0.005	0.89	0.86/ 0.79
	Removing Item 5 and 6	0.043	0.004	0.005	0.90	0.85/ 0.79
	Removing Occasion 1	0.037	0.004	0.005	0.87	0.87/ 0.81
	Removing Occasion 2	0.029	0.006	0.005	0.85	0.85/ 0.72
	Removing Occasion 3	0.037	0.007	0.007	0.84	0.84/ 0.73
	Removing Occasion 4	0.043	0.003	0.007	0.86	0.86/ 0.81
	Removing Occasion 5	0.044	0.005	0.007	0.87	0.87/ 0.78

results, accommodating potential variations and promoting a transparent, holistic evaluation of the data at hand. Thus, substantial person variance suggests a predominant presence of enduring distress elements captured by K-10, consistent with the earlier assessments of the scale’s psychometric properties.

It is worth noting that distress in older populations can often be challenging to detect due to various factors including a general tendency

among older adults to underreport psychological distress and the overlapping symptoms of distress with other aging-related health issues. This reality might have influenced the minimal changes noted over time in the study, where the true dynamics of psychological distress could potentially be masked or understated. Recognizing this, it is essential to approach the interpretation of our results with a nuanced understanding that encompasses the complexity of distress manifestations in older individuals. Moreover, the substantial proportion of enduring distress depicted in our findings might, to an extent, reflect the aforementioned challenge in detecting nuanced shifts in distress levels over time in older adults.

The D-study further elucidated the enduring nature of specific symptoms like feelings of worthlessness and fatigue, spotlighting them as potential focal points for intervention strategies, considering their high Test-Criterion Index (TCI) values. Contrastingly, items reflecting “restlessness” exhibited less enduring patterns, proposing a viable entry point for early interventions. Moreover, our investigation aligns with findings from studies on similar scales (Miller et al., 2021; Paterson et al., 2018), affirming the prudence in retaining all original items to maintain the scale’s reliability. In addition, we found no difference in G-coefficients when removing one assessment occasion from the analyses at a time. This indicated that these results cannot be attributed to specific effects of a particular occasion.

We also noticed a significant uptick in distress scores in the period between the fourth and fifth occasions, possibly attributed to the lengthier interval and age-related factors. This observation coincides with Kumar et al. (2022) and Moulton et al. (2020), highlighting a trend of escalating distress with advancing age, albeit with a potential contribution from a segment of individuals with dementia. Future investigations might delve deeper into this aspect, exploring the interplay between psychological disorders and age-related distress dynamics.

5.1. Limitations and future directions

The study bears certain limitations, primarily emanating from the restricted demographic representation, hinting at a necessity for broader, more diverse samples to enhance the findings’ applicability (Pickett and Wilkinson, 2010). Additionally, the exclusion criteria, targeting a population with relatively stable mental health histories, potentially influenced the observed K-10 score stability, suggesting a venue for future studies encompassing a wider demographic spectrum to capture more nuanced distress patterns (Medvedev et al., 2017; Paterson et al., 2018).

6. Conclusions

To sum up, our research underscores the K-10 scale’s reliability in gaging enduring psychological distress in older adults, showcasing its

potential as a tool for evaluating intervention efficacy over time. Future studies might consider shorter time intervals between assessments to discern more precisely between dynamic and enduring distress patterns.

Funding

The MAS study was funded by three National Health & Medical Research Council (NHMRC) of Australia Program Grants (ID350833, ID568969, and APP1093083). <https://www.nhmrc.gov.au/funding>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

Compliance with ethical standards

The MAS was approved by the University of New South Wales Human Research Ethics Committee (HC 05,037, 09,382, 14,327) and all participants provided written consent to participate in the study.

Data availability statement

The data are deposited at the centre for Healthy Brain Ageing (CHeBA) Research Bank containing research data and biospecimens that have been collected during the course of research projects conducted at the University of New South Wales ABN 57 195 873 179, a body corporate established pursuant to the University of New South Wales Act 1989 (NSW) of UNSW Sydney NSW 2052, Australia.

Preregistration

This study is pre-registered at AsPredicted and can be accessed on the following link <https://aspredicted.org/~Xi6BkmTqpF>.

CRediT authorship contribution statement

Vivienne Yu-X Yan: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Katya Numbers:** Data curation, Project administration, Validation, Writing – review & editing. **Preminder S. Sachdev:** Funding acquisition, Investigation, Project administration, Writing – review & editing. **Henry Brodaty:** Data curation, Funding acquisition, Project administration, Writing – review & editing. **Oleg N. Medvedev:** Conceptualization, Methodology, Software, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that there is no conflict of interest.

Acknowledgements

We thank the participants and their informants for their time and generosity in contributing to this research. We also acknowledge the MAS research team: <https://cheba.unsw.edu.au/research-projects/sydney-memory-and-ageing-study>

References

- Arterberry, B.J., Martens, M.P., Cadigan, J.M., Rohrer, D., 2014. Application of generalizability theory to the big five inventory. *Pers. Individ. Dif.* 69, 98–103. <https://doi.org/10.1016/j.paid.2014.05.015>.
- Arnaud, Malet, L., Teissedre, F., Izaute, M., Moustafa, F., Geneste, J., Schmidt, J., Llorca, P.-M., Brousse, G., 2010. Validity study of Kessler's psychological distress scales conducted among patients admitted to french emergency department for alcohol consumption-related disorders. *Alcohol. Clinic. Exp. Res.* 34 (7), 1235–1245. <https://doi.org/10.1111/j.1530-0277.2010.01201.x>.
- Australian Bureau of Statistics, 2022. *Census of Population and Housing: Population Data Summary, 2021*. <https://www.abs.gov.au/statistics/people/population/population-census/latest-release>.

- 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. *J. Nerv. Ment. Dis.* 198 (9), 660–664. <https://doi.org/10.1097/nmd.0b013e3181ef1f16>.
- Brooks, R.T., Beard, J., Steel, Z., 2006. Factor structure and interpretation of the K10. *Psychol. Assess.* 18 (1), 62. <https://doi.org/10.1037/1040-3590.18.1.62>.
- Cardinet, J., Johnson, S., Pini, G., 2011. *Applying Generalizability Theory Using EduG*. Routledge.
- Chalmers, R.A., Pratscher, S.D., Bettencourt, B., Medvedev, O.N., 2021. Applying generalizability theory to differentiate between trait and state in the interpersonal mindfulness scale (IMS). *Mindfulness* 12 (3), 613–622. <https://doi.org/10.1007/s12671-020-01520-5>.
- Conner, Copeland, V.C., Grote, N.K., Koeske, G., Rosen, D., Reynolds, C.F., Brown, C., 2010. Mental health treatment seeking among older adults with depression: the impact of stigma and race. *Am. J. Geriatr. Psychiatry* 18 (6), 531–543. <https://doi.org/10.1097/JGP.0b013e3181cc0366>.
- Cronbach, L.J., Rajaratnam, N., Gleser, G.C., 1963. Theory of generalizability: a liberalization of reliability theory. *Br. J. Stat. Psychol.* 16 (2), 137–163. <https://doi.org/10.1111/j.2044-8317.1963.tb00206.x>.
- Doris, S.F., Lee, D.T., Woo, J., Thompson, D.R., 2004. Correlates of psychological distress in elderly patients with congestive heart failure. *J. Psychosom. Res.* 57 (6), 573–581. <https://doi.org/10.1016/j.jpsychores.2004.04.368>.
- Drapeau, A., Marchand, A., Beaulieu-Prévost, D., 2012. Epidemiology of psychological distress. *Mental Illnesses-Understand. Predict. Control* 69 (2), 105–106. <https://doi.org/10.5772/30872>.
- Easton, S.D., Safadi, N.S., Wang, Y., Hasson, R.G., 2017. The Kessler psychological distress scale: translation and validation of an Arabic version. *Health Qual. Life Outcomes* 15 (1), 1–7. <https://doi.org/10.1186/s12955-017-0783-9>.
- Forrest, Siegert, R.J., Krägeloh, C.U., Landon, J., Medvedev, O.N., 2021. Generalizability theory distinguishes between state and trait anxiety. *Psychol. Assess.* 33 (11), 1080–1088. <https://doi.org/10.1037/pas0001060>.
- Hobbs, M.J., Joubert, A., Mahoney, A.E.J., Andrews, G., 2018. Treating late-life depression: comparing the effects of Internet-delivered cognitive behavior therapy across the adult lifespan. *J. Affect. Disord.* 226 (2), 58–65. <https://doi.org/10.1016/j.jad.2017.09.026>.
- Huisman, 2000. Imputation of missing item responses: some simple techniques. *Qual. Quant.* 34 (4), 331–351. <https://doi.org/10.1023/A:1004782230065>.
- Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., 2002. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol. Med.* 32, 959–976. <https://doi.org/10.1017/S0033291702006074>.
- Kessler, R.C., Barker, P.R., Colpe, L.J., Epstein, J.F., Gfroerer, J.C., Hiripi, E., Howes, M. J., Normand, S.L., Manderscheid, R.W., Walters, E.E., Zaslavsky, A.M., 2003. Screening for serious mental illness in the general population. *Arch. Gen. Psychiatry* 60 (2), 184–189. <https://doi.org/10.1001/archpsyc.60.2.184>.
- Kumar, S.S., Merkin, A.G., Numbers, K., Sachdev, P.S., Brodaty, H., Kochan, N.A., Trollor, J.N., Mahon, S., Medvedev, O., 2022. A novel approach to investigate depression symptoms in the aging population using generalizability theory. *Psychol. Assess.* <https://doi.org/10.1037/pas0001129>.
- Lace, Greif, T.R., McGrath, A., Grant, A.F., Merz, Z.C., Teague, C.L., Handal, P.J., 2019. Investigating the factor structure of the K10 and identifying cutoff scores denoting nonspecific psychological distress and need for treatment. *Mental Health Prevent.* 13, 100–106. <https://doi.org/10.1016/j.mhp.2019.01.008>.
- Lu, S.H., Dear, B.F., Johnston, L., Wootton, B.M., Titov, N., 2014. An internet survey of emotional health, treatment seeking and barriers to accessing mental health treatment among Chinese-speaking international students in Australia. *Couns. Psychol. Q.* 27 (1), 96–108. <https://doi.org/10.1080/09515070.2013.824408>.
- McShane, C.J., Quirk, F., Swinbourne, A., 2016. Development and validation of a work stressor scale for Australian farming families. *Aust. J. Rural Health* 24 (4), 238–245. <https://doi.org/10.1111/ajr.12261>.
- Medvedev, 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., Siegert, R.J., 2022. Generalizability theory. *Handbook of Assessment in Mindfulness Research*. Springer, Cham. https://doi.org/10.1007/978-3-030-77644-2_5-1.
- Medvedev, O., Truong, Q., Merkin, A., Borotkanics, R., Krishnamurthi, R., Feigin, V., 2021. Cross-cultural validation of the stroke riskometer using generalizability theory. *Sci. Rep.* 11 (1), 1–10. <https://doi.org/10.1038/s41598-021-98591-8>.
- Merson, Newby, J., Shires, A., Millard, M., Mahoney, A., 2021. The temporal stability of the Kessler psychological distress scale. *Aust. Psychol.* 56 (1), 38–45. <https://doi.org/10.1080/00050067.2021.1893603>.
- Miller, 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. *Int. J. Stress Manag.* 28 (2), 147–153. <https://doi.org/10.1037/str0000207>.
- Moult, Kingstone, T., Chew-Graham, C.A., 2020. How do older adults understand and manage distress? A qualitative study. *BMC Fam. Pract.* 21 (1), 77. <https://doi.org/10.1186/s12875-020-01152-7>.
- Nurse Key, 2022. *Mental Disorders of Old Age*. <https://nursekey.com/mental-disorders-of-old-age/>.
- Paterson, J., Medvedev, O.N., Sumich, A., Tautolo, E.S., Krägeloh, C.U., Sisk, R., Siegert, R.J., 2018. Distinguishing transient versus stable aspects of depression in New Zealand Pacific Island children using generalizability theory. *J. Affect. Disord.* 227, 698–704. <https://doi.org/10.1016/j.jad.2017.11.075>.

- Paúl, R.O., Santos, P., 2009. Cognitive impairment in old people living in the community. *Arch. Gerontol. Geriatr.* 51 (2), 121–124. <https://doi.org/10.1016/j.archger.2009.09.037>.
- Pickett, Wilkinson, R.G., 2010. Inequality: an underacknowledged source of mental illness and distress. *Br. J. Psychiatry* 197 (6), 426–428. <https://doi.org/10.1192/bjp.bp.109.072066>.
- Ridner, 2004. Psychological distress: concept analysis. *J. Adv. Nurs.* 45 (5), 536–545. <https://doi.org/10.1046/j.1365-2648.2003.02938.x>.
- Sachdev, Brodaty, H., Reppermund, S., Kochan, N.A., Trollor, J.N., Draper, B., Slavin, M. J., Crawford, J., Kang, K., Broe, G.A., Mather, K.A., Lux, O., 2010. The Sydney Memory and Ageing Study (MAS): methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70–90 years. *Int. Psychogeriatr.* 22 (8), 1248–1264. <https://doi.org/10.1017/S1041610210001067>.
- Shavelson, Webb, N.M., Rowley, G.L., 1989. Generalizability theory. *Am. Psychol.* 44 (6), 922–932. <https://doi.org/10.1037/0003-066X.44.6.922>.
- Smout, 2019. The factor structure and predictive validity of the Kessler Psychological Distress Scale (K10) in children and adolescents. *Aust. Psychol.* 54 (2), 102–113. <https://doi.org/10.1111/ap.12376>.
- Swiss Society for Research in Education Working Group, 2006. *EDUG User Guide*. IRDP, Eucatel, Switzerland.
- Thelin, Mikkelsen, B., Laier, G., Turgut, L., Henriksen, B., Olsen, L.R., Larsen, J.K., Arnfred, S., 2017. Danish translation and validation of Kessler's 10-item psychological distress scale - K10. *Nord. J. Psychiatry* 71 (6), 411–416. <https://doi.org/10.1080/08039488.2017.1312517>.
- Truong, 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* 11 (4), 953–963. <https://doi.org/10.1007/s12671-020-01324-7>.
- Vispoel, W.P., Morris, C.A., Kilinc, M., 2018. Applications of generalizability theory and their relations to classical test theory and structural equation modeling. *Psychol. Methods* 23 (1), 1–26. <https://doi.org/10.1037/met0000107>.
- Voyer, Verreault, R., Cappeliez, P., Holmes, D., Mengue, P.N., 2005. Symptoms of psychological distress among older adults in Canadian long-term care centres. *Aging Ment. Health* 9 (6), 542–554. <https://doi.org/10.1080/13607860500193336>.
- Weissman, J.S., Pratt, L.A., Miller, E.A., Parker, J.D., 2015. *Serious Psychological Distress Among Adults, United States, 2009-2013*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.
- Welsh, J., Korda, R.J., Banks, E., Strazdins, L., Joshy, G., Butterworth, P., 2020. Identifying long-term psychological distress from single measures: evidence from a nationally representative longitudinal survey of the Australian population. *BMC Med. Res. Methodol.* 20 (1), 1–9. <https://doi.org/10.1186/s12874-020-00938-8>.
- World Health Organization (WHO), 2022. *Aging and Health*. World Health Organization (WHO). <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>.