Running head: Age-related differences in psychopathology

# Age-related differences in internalizing psychopathology amongst the Australian general population

Matthew Sunderland[1]

Tim Slade[1]

Natacha Carragher[1]

Philip Batterham[2]

Heather Buchan[1]

***Manuscript in preparation for Journal of Abnormal Psychology***

[1] NHMRC Centre for Research Excellence in Mental Health and Substance Use, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, New South Wales, Australia.

[2] Centre for Mental Health Research, the Australian National University, Canberra, Australian Capital Territory, Australia.

## ABSTRACT

Two methodological criticisms have limited the reliability and validity of findings from previous studies that seek to examine change across the lifespan in levels of internalizing psychopathology using general population surveys. The first criticism involves the potential influence of cohort effects that confound true age-related changes while the second criticism involves the use of a single form of assessment to measure and compare levels of internalizing psychopathology. This study seeks to address these criticisms by modelling age-related change using multiple measures and multiple surveys. Data from two epidemiological surveys conducted ten years apart in the Australian general population were combined and utilized for the current study. The latent construct of internalizing psychopathology was modelled using a combination of DSM-IV depression and anxiety diagnoses as well as items from the Kessler Psychological Distress scale (K10; Kessler et al., 2002). Confirmatory factor analysis indicated that a single internalizing dimension provided good model fit to the data. Multi-group confirmatory factor analysis indicated that strict measurement invariance of the model can be assumed across survey administrations and age bands, justifying comparisons of mean differences in latent trait levels. Significant changes in mean levels of latent internalizing psychopathology were evident between respondents aged 30-39 years old in 1997 and respondents aged 40-49 years old in 2007, suggesting a minor but significant increase in psychopathology across middle age. By contrast, a minor but significant decrease in psychopathology was noted when transitioning from late middle age (50-59 years old) to old age (60-69 years old). The majority of individuals in the general population will experience constant levels of internalizing psychopathology as they age, suggesting that the construct is relatively stable.


**Keywords: prevalence, psychopathology, age comparisons, bias, internalizing, DSM-IV**

## INTRODUCTION

Accurately estimating the prevalence of internalizing disorders (unipolar mood and anxiety disorders) in the general population as well as age-related changes is critical for improving health service planning and identifying areas of scarce resources (Thomas, 1998). As members of the post-World War II population boom grow older they may experience increased levels of psychopathology, flagging the need to future-proof mental health care services to ensure they can cope with a possible influx of patients with mental health issues. Alternatively, if mental disorders are shown to decline with age then it may be premature to assign extra resources to old age mental health services at the expense of health services dedicated to addressing chronic physical conditions. In a similar fashion, clinicians need to better understand age-related changes in mental health so that they know the likelihood of a particular disorder in a patient of a given age. For example, if depression is found to decline with age, then clinicians may neglect to attribute various symptoms to depression amongst older patients in comparison to younger patients (Snowdon, 2001).

Turning to the extant literature, epidemiological studies conducted in multiple countries have previously demonstrated that prevalence rates of internalizing disorders differ substantially across the lifespan. Jorm (2000) compiled a review of psychiatric epidemiological studies and observed a common trend of depression and anxiety steadily increasing across young and middle-age followed by a drop in old age, although there was no consensus regarding the peak age of prevalence. More recent epidemiological evidence from the United States and the World Mental Health Survey Initiative confirmed these findings and indicated that a significant decrease in the prevalence of major depression amongst older adults was evident in seven out of 10 developed countries (Belgium, France, Germany, Japan, Netherlands, New Zealand, and the United States) (Kessler et al., 2010a; Kessler et al., 2010b). In Australia, the lower prevalence of mental disorders experienced by older adults in

comparison to middle-aged adults is consistent with the majority of developing countries (Trollor, Anderson, Sachdev, Brodaty, & Andrews, 2007). Additionally, the prevalence of any anxiety disorder remained relatively constant across young and middle age before demonstrating a steep decrease from 20.4% in those aged 55-64 years to 12.6% in those aged 65-74 years (McEvoy, Grove, & Slade, 2011).

The finding of significant age-related differences in internalizing disorders across the lifespan from epidemiological surveys has been met with substantial scepticism in the literature, primarily due to the influence of cohort effects inherent in single cross-sectional studies. The comparison of birth cohorts in single cross-sectional studies can easily mask true differences with age (Jorm, 2000; Lewinsohn, Rohde, Seeley, & Fischer, 1993). For example, it is possible that younger compared to older adults may differentially exposed to various environmental and situational factors that contribute to the emergence of mental health issues across the lifespan (e.g. better recognition and cultural attitudes toward mental health in more recent cohorts) (Simpson, Meadows, Frances, & Patten, 2013). Therefore differences in mental health prevalence between younger and older adults could simply be due to differences in the exposure level to environmental or situational risk factors rather than true differences occurring with age. Methods to remove the influence of cohort effects primarily involve the collection of longitudinal data over many years.

One recent longitudinal study of probable depression and depressive symptomology followed 35,200 community-living participants aged 45-103 years at baseline for 13 years and found that the likelihood of probable depression remained constant. However, there was some indication that the reporting of depressive symptomatology increased over the 13 year period in people aged 70+ years at baseline (Burns, Butterworth, Luszcz, & Anstey, 2013). Whilst this study should be commended for utilizing a large pooled dataset to examine age-related change, the study was limited in the sense that it could not examine a larger range of

DSM-IV internalizing disorders, nor did it have the capacity to provide information on age-related changes amongst community members in the younger age ranges (e.g. 18-44 years). In fact, population-based longitudinal studies examining age-related changes in internalizing disorder prevalence are sorely lacking in the extant literature, most likely due to excessive research costs, time constraints, and high respondent attrition levels.

Other, more cost-effective methods to remove the influence of cohort effects from age-related changes in internalizing psychopathology involve the collection of repeated cross-sectional surveys. Using multiple surveys facilitates the comparison of younger and older adults born within the same time period (Smith, 2008). Consider, for instance, two different groups of people, one interviewed in the year 1990 when they were 20 years of age, and another interviewed in the year 2000 when they were 30 years of age. Both groups of people were born in the same year (1970) and thus can be considered to come from the same birth cohort. Effectively the same early environmental and situational factors affecting the experience of internalizing psychopathology should apply equally to both groups. In this way, any observed differences between the two age groups can be better accounted for by the aging of ten years in comparison to methods that compare different cohorts within a single cross-sectional survey. That being said, this method cannot completely rule out the possibility that the mental health status of the older group may be influenced by the current cultural and historical context. Using our above example, the mental health status of the 30 year olds measured in 2000 could have been influenced by the aging of ten years as well as any significant cultural or historical changes that occurred within that ten year period between 1990 and 2000. Previous studies utilizing this method have generally indicated that prevalence estimates and levels of psychopathology remain relatively consistent and stable across time amongst individuals belonging to the same birth cohort (Brault, Meuleman, & Bracke, 2012; Fu, Lee, Gunnell, Lee, & Cheng, 2013; Spiers et al., 2011).

A second problem faced by the previous literature, with respect to psychiatric epidemiological studies that utilize DSM-IV and ICD-10 criteria to determine prevalence, is the rising concern of age-related measurement bias. O'Connor and Parslow (2009) reasoned that long and complex questions contained in epidemiological surveys require analyses of multiple time frames and attributions, something that may be particularly difficult amongst older adults whose attention span, working memory, and information processing speed are typically reduced. As a result of the complexities required with responding, many older participants may automatically respond negatively, as this tends to be the easiest response, or search for a more plausible explanation, such as physical attributions to psychological symptoms (Knauper & Wittchen, 1994). O'Connor and Parslow (2009, 2010a) provided some evidence for this argument by demonstrating that positive responses to complex questions in the Composite International Diagnostic Interview (CIDI) tend to decline with age at a steeper rate than shorter, simple questions contained in the Kessler 10 psychological distress scale (K10), which theoretically measures similar constructs. In addition to age-related bias, there is also the concern that comparisons of individual differences may be increasingly influenced by measurement errors inherent in the use of a single instrument to derive a diagnosis. For instance vague or overly complicated diagnostic questions, problems with recall or a lack of introspection, inconsistent or confusing time frames or response categories, and sensitive content that may induce socially desirable responding patterns.

One method to reduce the level of measurement bias attributed to the use of a single instrument when assessing global levels of mental health is to use multiple instruments that are merged into a composite form of assessment. For instance, multiple instruments that are deemed to assess the same construct can be combined by forming a latent variable that explains the variance between all manifest items from multiple instruments. This has the advantage of counterbalancing any confounding and negative effects of each instrument with

the strengths of multiple assessments (e.g. reducing the amount of error). Change of global levels of mental health across age amongst individuals can then be examined using the more reliable and robustly measured latent variable rather than estimates that are derived from a single form of assessment. That being said, group comparisons of mental health may still be susceptible to levels of bias despite using a composite form of measurement, particularly if one group of individuals is uniformly affected by the biasing factor in comparison to the other group, as would be the case with age-related bias discussed above. For example, one age group might respond to the items across the multiple forms of measurement more consistently than another, resulting in changes of the covariance matrix across age groups. Therefore, the latent variable should be constructed using items that are relatively robust against bias that affect whole groups of individuals as well as testing more restrictive levels of measurement invariance between age groups prior to comparing latent means (Vandenberg & Lance, 2000).

The internalizing latent variable, represented by multiple forms of measurement, could represent an ideal candidate to measure age-related change in latent levels of depression and anxiety. This latent variable could be constructed by combining the assessment of DSM-IV diagnoses, such as the mood and anxiety disorders, with the K10 items that represent non-specific psychological distress. Previously the internalizing latent variable has been predominately constructed using the assessment of DSM-IV diagnoses, including major depression, generalized anxiety disorder, panic disorder, and social phobia, which have been assessed using comprehensive questions in epidemiological studies (Krueger, 1999; Krueger & Markon, 2006; Slade & Watson, 2006). Previous evidence using waves of longitudinal data has also demonstrated that this internalizing latent variable is relatively stable across age (Eaton, Krueger, & Oltmanns, 2011; Eaton et al., 2013; Krueger, Caspi, Moffitt, & Silva, 1998; Vollebergh et al., 2001). Likewise, the K10 scale has

demonstrated a strong relationship with DSM-IV diagnoses of internalizing disorders and therefore may further explain global levels of internalizing psychopathology (Slade, Grove, & Burgess, 2011; Sunderland, Slade, Stewart, & Andrews, 2011). Importantly, the K10 items have been shown to be robust to age-related bias when comparing differences between age groups (O'Connor & Parslow, 2010b; Sunderland, Hobbs, Anderson, & Andrews, 2012). Therefore any latent measure of internalizing that combines the comprehensive assessment of DSM-IV disorders with the K10 items is likely to be more robust than the use of any single form of assessment.

Given the significant improvements that can be attained by using the above mentioned methods, this paper has two main foci. The first aim is to construct a latent variable of internalizing psychopathology using both DSM-IV diagnoses and K10 items, thereby reducing the level of bias associated with using a single form of measurement. The second aim is to subsequently use the latent variable of internalizing psychopathology and compare groups of respondents who were born in the same birth cohorts using epidemiological surveys administered in 1997 and in 2007. Any observed changes across the ten year time period within the same birth cohorts can be confidently interpreted without significant levels of measurement and age-related bias or the influence of comparing different birth cohorts, which have proven problematic in previous single-study papers that make age-related comparisons using epidemiological data.

## METHODS

### Sample

Data for the current study were derived from the 1997 and 2007 Australian National Surveys of Mental Health and Wellbeing (NSMHWB). The two surveys were conducted under the auspices of the Australian Bureau of Statistics (ABS) and administered face-to-face by trained lay-interviewers. The surveys generated representative estimates of non-

institutionalized Australian private households (excluding remote or sparsely populated areas) using a stratified, multi-stage area sampling design (survey response rate of 78% in 1997 and 60% in 2007). In all subsequent analyses these data were weighted for the sex and age distribution of the Australian general population according to the most recent census data at the time of interview administration. The 1997 survey randomly sampled one member of each household aged 18 years or older whilst the 2007 survey randomly sampled one member of each household aged 16 years or older with an upper limit of 85 years. In 2007, the younger (16-24 years) and older (65-85 years) age groups were over-sampled to ensure reliable estimates amongst these traditionally under-represented age categories. More information regarding the methods, design, and sampling of the 1997 and 2007 NSMHWB as well as a detailed description of the respondent characteristics have been described previously (see Andrews, Henderson, & Hall, 2001; Slade, Johnston, Oakley Browne, Andrews & Whiteford, 2009). Given the age limit differences between the two surveys, the sample was restricted to a lower age limit of 20 years and an upper age limit of 69 years as a means to reliably compare matching ten year age cohorts. This resulted in an analyzed sample of 9,159 respondents in the 1997 survey and 6,865 respondents in the 2007 survey, making a combined sample size of 16,024.

**Measures**

*DSM-IV Internalizing disorders*

To measure DSM-IV internalizing disorders, the 1997 and 2007 surveys utilized the Composite International Diagnostic Interview version 2.1 (CIDI 2.1; Andrews & Peters, 1998; World Health Organization, 1997) and the World Mental Health version of the Composite International Diagnostic Interview (WMH-CIDI; Kessler & Ustun, 2004), respectively. Both interviews have been used extensively for psychiatric epidemiological studies and part of the World Mental Health Survey initiative, consisting of at least 28

countries representing all regions of the world. Both interviews possess sound psychometric properties and have been clinically calibrated/validated using gold standard semi-structured psychiatric interviews (Jordanova, Wickramesinghe, Gerada, & Prince, 2004; Kessler et al., 2004). Modifications to the WMH-CIDI were made for the specific purpose of the 2007 NSMHWB under the supervision of senior academics and survey experts. Most notably, the post-traumatic stress disorder (PTSD) module was edited to remove symptom questions related to the respondent's worst traumatic event. Both surveys assessed DSM-IV criteria for depression, dysthymia, mania, panic disorder, social phobia, agoraphobia, generalized anxiety disorder, substance use disorders, post-traumatic stress disorder, and obsessive compulsive disorder.

DSM-IV diagnostic information was obtained with reference to the respondent's lifetime in 2007 whereas 1997 utilized past twelve months point prevalence. In 2007, as a means of aiding respondent recall a series of prompts and techniques were used when attempting to elicit precise information, such as age of onset. In addition, the interview sought to acquire information on the respondent's worst episode to aid recall of symptoms and form a diagnosis on the most salient event across the respondent's lifetime. Additional recency questions were included to determine whether lifetime disorder symptoms had been present in the 12 months and 30 days prior to interview. In the 1997 survey, the recency question was utilized to determine whether symptoms experienced in the past 12 months had been present in the 30 days prior to interview. A confirmed lifetime or 12 month diagnosis, plus the presence of symptoms in the past 30 days, were considered sufficient to generate a 30 day point prevalence estimate for each DSM-IV disorder. The 30 day point prevalence data from both the 1997 and 2007 surveys were utilized in the remaining analysis to coincide with the 30 day time frame of the K10.

Only disorders that have reliably and robustly demonstrated to load significantly on the internalizing construct were utilized for the current study (Krueger, 1999; Slade & Watson, 2006). Disorders present in the past 30 days that formed the internalizing construct in the current study included: major depressive disorder, social phobia, generalized anxiety disorder, post-traumatic stress disorder, and panic disorder with or without agoraphobia. Dysthymia was excluded from the model given the very low prevalence of the disorder observed in the past 30 days amongst some of the age groups.

*Non-specific Psychological Distress*

Non-specific psychological distress was measured in both surveys using the K10. This scale was originally designed as a short screening instrument to measure psychological distress attributed to depression and anxiety experienced in the past 30 days. Recently, studies have demonstrated that the K10 can be used as an effective and efficient screening tool for DSM-IV internalizing disorders (Kessler et al., 2010c; Sunderland, Slade, Stewart, & Andrews, 2011). The K10 has very strong psychometric properties and has demonstrated good reliability and validity across multiple settings and populations (Arnaud et al., 2010; Baillie, 2005; Donker et al., 2009; Furukawa et al., 2008; Hides et al., 2007; Andrews & Slade, 2001; Spies et al., 2009). Previously, the K10 has demonstrated a factor structure consisting of one latent factor, representing psychological distress, in general population samples (Sunderland, Mahoney, & Andrews, 2012).

Each of the 10 items is designed to assess the presence of symptoms attributed to depression and anxiety in the past 30 days. These items include: feeling tired, nervous, so nervous that nothing could calm you down, hopeless, restless or fidgety, so restless that you could not sit still, depressed, so depressed that nothing could cheer you up, that everything was an effort, and worthlessness. Each item is assessed on a 5-point Likert-type scale that ranges from ‘did not experience the symptom at all’ to ‘experienced the symptom all of the

time in the past 30 days'. Traditionally, surveys conducted in Australia score the K10 by summing the 10 items to generate a total score between 10 and 50, with higher scores indicating higher levels of psychological distress. Normative data from the Australian general population indicates that the majority of individuals receive a mean K10 score between 12.6 and 15.0 (Slade, Grove, & Burgess, 2011).

**Statistical Analysis**

The analyses planned for the current study were broken down into three iterative stages. The first stage involved investigating the factorial structure of internalizing psychopathology using DSM-IV internalizing disorders present in the past 30 days and non-specific psychological distress by means of exploratory and confirmatory factor analysis. The structure was examined separately in both surveys before the samples were merged and the structure was replicated once again. This stage enabled confirmation that the DSM-IV disorders and K10 items could be reliably combined to form a latent variable representing internalizing psychopathology. The second stage, involved the establishment of measurement invariance of the internalizing construct across survey administrations and across ten year age groups using multi-group confirmatory factor analysis. The establishment of measurement invariance facilitates comparisons across age in the internalizing latent variable and was used as further justification for combining the two surveys to measure change across the same birth cohort. The final stage involved a comparison of latent means between similar cohorts ten years apart in order to determine the shifts of internalizing psychopathology across a ten year period in the same age cohorts. All analyses were conducted using Mplus version 7 (Muthén & Muthén, 2010). Item responses were weighted for the Australian general population. However, due to software restrictions, the standard errors were not adjusted to take into account the complex sampling design of the 1997 and 2007 NSMHWB. Instead, the standard errors were treated as if they were derived from a simple random sample.

# RESULTS

## Exploratory and Confirmatory Factor Analysis

Exploratory and confirmatory factor analyses were conducted separately for each survey and on the combined survey dataset using the Weighted Least Squares Mean and Variance adjusted (WLSMV) estimator, which is well suited to analyses of categorical and ordinal data (Muthén & Muthén, 2010). However, given that the separate analyses provided almost identical conclusions only the results for the merged sample are provided below. The smallest, and most parsimonious, number of factors to explain variation amongst the internalizing disorders and K10 items was determined using the eigenvalues and scree plot method. The EFA generated a scree plot with a sharp decrease in eigenvalues from the first to the second factor solutions (8.71 to 1.02). This provided sufficient evidence that most of the variance between the indicator variables could be explained by a one-factor solution reflecting internalizing psychopathology. The remaining factor solutions evidenced eigenvalues at or below 1.0. The one-factor solution provided good model fit (CFI = 0.94, RMSEA = 0.057, RMSEA 90% CI = 0.055-0.058) according to previously defined criteria suggesting that CFI scores closer to 1.00 and RMSEA scores lower than 0.08 indicate good model fit (Browne & Cudeck, 1993; Hu & Bentler, 1998). In addition, inspection of the Quartimin rotated factor loadings for the one factor model revealed that each item was significant at the 0.05 level and salient (i.e. a factor loading greater than 0.40). Closer inspection of the two factor solution revealed that only two items loaded onto the second factor, representing K10 items “restless” (0.898) and “so restless” (0.874). It was concluded that the second factor solution may be detecting the additional item dependency between the K10 items since respondents were not queried about the “so restless” item if they did not previously endorse feeling “restless” in the past 30 days. The best fitting factor solution and factor loadings are provided in Figure 1.

Five CFA models, informed by the EFA and extant literature (Krueger, 1999; Slade & Watson, 2006; Sunderland, Mahoney, & Andrews, 2012), were examined and compared using the merged data: **Model A** (a one-factor model), **Model B** (a one-factor model with correlated errors between K10 items "depressed" and "so depressed", "restless" and "so restless", and "nervous" and "so nervous", as a means to take into account the automatic skip instructions that contribute to common variance between these items not explained by the latent factor [Sunderland, Mahoney, & Andrews, 2012]), **Model C** (a two-factor model with all DSM-IV disorders loading on one factor [disorder] and all K10 items loading on another [distress]), **Model D** (a two-factor model taking into account the common variance between related K10 items), and **Model E** (a two-factor model with factor covariances fixed to 0). To set the scale and identify the measurement model, the factor loading for the first item on each separate factor were constrained to 1.0 by default. Previous evidence has indicated that the internalizing latent construct could be modeled using a hierarchical structure with two lower order sub-factors representing fear and distress disorders (Krueger, 1999; Slade & Watson, 2006). However for the current study the models were limited to the single internalizing factor without the sub-factors to better coincide with the single distress factor that represents the K10 items (Sunderland, Mahoney, & Andrews, 2012).

The fit indices (CFI and RMSEA values) for each of the five models in the merged dataset are provided in Table 1. Model B and Model D provided very similar CFI and RMSEA values demonstrating excellent model fit. Examination of the factor covariance between the two factors in Model D, however, indicated some degree of multicollinearity as evidenced by a very high factor covariance of 0.912. The high correlation implies that both factors are effectively tapping the same underlying construct and therefore Model B was selected as the best fitting model on the grounds of parsimony. These results suggest that

psychological distress measured by the K10 and DSM-IV disorder rates can be used interchangeably to predict the underlying level of latent internalizing psychopathology.

**Measurement Invariance testing**

Measurement invariance was first determined between the two survey administrations (group 1 = 1997 survey, group 2 = 2007 survey) by comparing a series of nested multi-group confirmatory factor models. The analysis proceeded in a series of steps by fitting nested models that estimate increasing levels of measurement invariance (Muthén & Asparouhov, 2002). First, a baseline model was established by assuming configural invariance of the best fitting model. This was achieved by imposing the same factor structure across each of the survey administrations and age groups but allowing the factor loadings, thresholds, and residual covariances to be freely estimated. In order to identify the model, the factor means for each group were constrained to 0 whilst the factor variances were constrained to 1.0. All residual variances were fixed to 1.0. Model fit for the baseline model was again determined using the CFI and RMSEA using the same guidelines as stated in the CFA.

Second, to assess scalar invariance, a model was estimated that constrained loadings and thresholds to equality across groups (note: for categorical items the loadings and thresholds are constrained to equality in tandem as both parameters influence the response curves). Factor means, variances, and residual variances were freely estimated across all groups except the referent group (i.e. 20-29 years at 1997). Third, to assess strict invariance, a model was estimated that constrained loadings, threshold, and residual variances to equality across groups. Factor means and variances were freely estimated across all groups except the referent group. Fourth, a strict invariance model was then estimated that constrained the residual covariances to equality across groups, all other parameters remained the same as the strict invariance model.

To determine the level of measurement invariance present in the current sample, the nested models were compared to each previous model beginning with scalar invariance. Changes in the pragmatic fit indices were relied on to evaluate change in model fit. According to Cheung and Rensvold (1999), a difference in the CFI values less than 0.01 represents a trivial difference in model fit. Furthermore, 90% confidence intervals around the RMSEA values for each model were calculated and meaningful differences between the models were assessed by non-overlapping confidence intervals. The results of the measurement invariance tests comparing survey administrations are provided in Table 2. The ΔCFI between successive measurement invariance models revealed trivial differences in model fit with a difference ranging from no difference to -0.002, well below the suggested cut-point of 0.01. The overlapping 90% confidence intervals around the RMSEA values provided further support for the claim that there were no substantive differences in model fit. Given these results, the strict invariance model (equal loadings, thresholds, and residual variances) with equal residual covariances was selected as the best fitting model and provides adequate support that the latent construct was measured between the two surveys in a similar manner.

Measurement invariance testing then continued by examining the model fit of increasing degrees of measurement invariance across the age bands present in each survey (group 1 = 20-29 years in 1997, group 2 = 20-29 years in 2007, group 3 = 30-39 years in 1997, group 4 = 30-39 years in 2007, group 5 = 40-49 years in 1997, group 6 = 40-49 years in 2007, group 7 = 50-59 years in 1997, group 8 = 50-59 years in 2007, group 9 = 60-69 years in 1997, group 10 =60-69 years in 2007). Sample size and frequencies of the manifest variables for each of the birth cohorts are provided in the online supplementary material. As above, the strict invariance model (equal loadings, thresholds, and residual variances) with equal residual covariances was selected as the best fitting model based on trivial differences in the

CFI values and overlapping 90% RMSEA confidence intervals. Given the presence of strict measurement invariance, the latent means can be compared across age groups from both surveys without any significant influence of measurement bias. Model fit indices for each of the nested measurement invariance models across age groups are provided in Table 3.

**Comparison of latent means**

Given the determination of measurement invariance, the final best fitting multi-group invariant factor model (strict invariant model with fixed residuals) was then estimated and used to compare group latent variances and means. Prior to comparing latent means, the assumption of equal latent variances was tested in order to facilitate interpretation of latent mean comparisons. If the latent variances can be assumed as equivalent and fixed to 1.0 across each group, this allows for the mean differences between groups to be interpreted as Cohen's *d* effect sizes. Testing of this assumption was achieved by fitting an additional multi-group CFA model of strict measurement invariance; however the latent variances across all groups were fixed to 1.0. Model fit was then compared to the strict measurement invariance model that freely estimated the latent variances. The difference between the two models in terms of CFI values (ΔCFI = -0.001) and overlapping 90% confidence intervals of the RMSEA, as shown in Table 3, indicated that the assumption of equal latent variances was upheld. A final multi-group CFA model was then tested that constrained the variances to 1.0 across each group as well as fixing the latent means to 0.0 across each group and comparing the fit to a model that freely estimated the means. In effect, this model demonstrates the level of overall significance regarding mean differences in the latent internalizing variable across all the age groups. As shown in Table 3, the change in model fit statistics indicated that there was no substantial reduction in model fit and the assumption of equal factor variances and means across the age groups was upheld.

Despite the above finding of equality of latent means, it is possible that minor but significant pairwise differences in the latent means exist across age groups. Therefore, it was decided to further investigate any pairwise differences to provide a more comprehensive picture of changing latent means. A series of pairwise comparisons between age groups were conducted in order to compare 10-year age differences in the same birth cohorts. For example, respondents who were born in 1968 through to 1977 were compared across the two surveys, resulting in a comparison between the same birth cohort at ages 20-29 and 30-39. Since Mplus does not model absolute values, latent means are compared by treating one group as a referent group and fixing the mean for that group to zero (as a method of identifying the latent scale). The latent mean values for the remaining groups can then be interpreted as the relative mean difference or mean deviation between the examined group and the referent group.

Pairwise mean deviation scores and significance tests for each of the birth cohorts are provided in Table 4. Latent mean levels of internalizing psychopathology tend to remain stable across the lifespan. However, there was a minor but significant increase ($d$=0.13, $p$=0.03) in latent levels of psychopathology across 10 years amongst the 1958-1967 birth cohort, suggesting an increase in internalizing problems during the transition from the thirties to the forties. Finally, there was a minor but significant decrease ($d$=-0.18, $p$=0.01) in latent levels of psychopathology across 10 years amongst the 1938-1947 birth cohort, suggesting that internalizing problems tend to recede when transitioning from the fifties to the sixties. There is a lack consensus regarding whether strict invariance is required to compare factor means in comparison to scalar invariance. To clarify this issue, the above mentioned mean comparisons were repeated using the scalar invariance model and the use of this model did not substantively alter the results and the conclusions.

## DISCUSSION

The current study sought to minimize the level of measurement bias when examining age differences in the level of internalizing psychopathology experienced across the lifespan. This was achieved by combining two forms of measurement that similarly tap internalizing psychopathology. Secondly, this study compared latent internalizing psychopathology across a 10-year time period in the same birth cohort by utilizing two epidemiological surveys conducted 10 years apart. The validity of comparing latent means across age groups and across multiple survey administrations was examined by testing increasingly restrictive levels of measurement invariance. The relative good fit of a model with strict measurement invariance confirmed that each age group and different survey administration were robust against bias attributed to the differential measurement of the latent construct. This finding also confirms that any systematic differences in how the latent variable indicators (i.e. DSM-IV diagnoses) were measured between the two surveys did not have a substantial impact on the measurement of the broad internalizing latent variable, enabling data from the two surveys to be merged into a common model.

Comparisons of latent means evidenced a degree of consistency in the mean level of internalizing psychopathology experienced across a 10-year age period within the same birth cohort. Given the invariant structure, we can infer then that any observed differences across age in DSM-IV disorders and non-specific psychological distress are due to differences in the latent internalizing construct. Indeed, there were two significant differences across the 10-year age period within two birth cohorts, demonstrating that internalizing psychopathology tends to peak during the transition into middle age whereas it decreases during the transition to older age. These findings offer some strength to the estimated age differences in DSM-IV and non-specific psychological distress found in previous epidemiological studies in the Australian population and worldwide (e.g. Kessler et al., 2010a; McEvoy, Grove & Slade, 2011; Trollor et al., 2009). That being said, the strength of the age differences within each

cohort, measured using Cohen's *d*, revealed that whilst the differences were significant they were relatively minor (*d*s ranging between 0.1-0.2), although the change may have a more important impact on a population and health services level. This finding is in contrast to the relatively large differences in prevalence found previously between age groups using single survey designs and supports the importance of removing the influence of cohort effects and the possibility of measurement bias in order to reduce over-inflated estimates of change (Brault, Meuleman, & Bracke, 2012). Moreover, this finding supports previous findings that examined age differences within birth cohorts from the 1993-2007 British Psychiatric Morbidity Surveys and found relatively stable levels of mental illness across age (Spiers et al., 2011).

The findings of the current study supports previous evidence that the latent structure of the common mental disorders (particularly the internalizing construct) remains stable across the lifespan and over time (Eaton, Krueger, & Oltmanns, 2011; Eaton et al., 2013; Krueger, Caspi, Moffitt, & Silva, 1998; Vollebergh et al., 2001). Indeed, Eaton, Krueger, & Oltmanns (2011) compared mean levels of internalizing psychopathology across ten years within birth cohorts and found moderate to high stability. They concluded that whilst individual DSM disorders may remit and recur over time, the underlying liability to develop and continue to experience these disorders remains relatively stable across age. The current results replicate and build on the results provided by Eaton, Krueger, & Oltmanns (2011) by examining a broader array of manifest indicators as well as more robust measurement of the internalizing latent construct. Consistent with their conclusion, these results provide further support for the use of dimensional latent variables in epidemiological and public health settings.

These findings are also consistent with the related argument that more genetically-based and biologically driven constructs, such as neuroticism and extraversion, remain

relatively stable across age with only minor downward shifts in late life, at least in comparison to some of the pronounced fluctuations observed in DSM-IV disorder prevalence across age (Allemand, Zimprich, & Hendriks, 2008; Johnson, McGue, & Krueger, 2005; Lucas & Donnellan, 2011; Mroczek & Spiro, 2003). Given that previous findings have established a strong correlation between the internalizing and neuroticism constructs (Griffith et al., 2010), it seems reasonable to assume that the internalizing construct is likewise a strong genetically/neurobiologically-based construct that is relatively stable across age and susceptible to minor downward shifts with increasing age. The implications of this include the possibility for further research to investigate the biological basis of internalizing psychopathology as well as the genetic and environmental foundations that result in either stability or fluctuations in internalizing psychopathology across time (Johnson, McGue, & Krueger, 2005).

The results of the current study must be considered in light of several limitations. First, in order to compare the same birth cohorts across a 10-year age period and given the differences in age limits exhibited by the two surveys, it was necessary to restrict the study to only those aged between 20 and 69 years, precluding our ability to examine any further age differences in the older (70+ years) age brackets. This has several implications given that previous evidence has demonstrated that whilst psychopathology decreases in old age, as partially demonstrated by the current study, it once again spikes during the oldest age (75+ years) bands as a result of advanced aging (Burns, Butterworth, Luszcz, & Anstey, 2013; Jokela, Batty, & Kivimaki, 2013; Snowdon, 2001). Therefore, we cannot assume that the downward trend in internalizing psychopathology demonstrated in those aged 60-69 years in the current study would continue into the more advanced age brackets.

Second, the study was only able to examine a difference within birth cohorts between 10 years measured across two time points. This limits further modeling of the trajectory and

course of internalizing psychopathology within specific cohorts. The relatively minor changes in internalizing psychopathology may be due to the fact that not enough time has passed within the cohorts to result in a significant change in mental health due to common aging factors. However, modeling change over extended time periods would necessitate multiple cohort studies conducted over many years or longitudinal data conducted on the same individuals across a lifetime. These studies are rare as they significantly increase financial costs and research burden. A third limitation involves the difference in survey response rates between the two survey administrations that may result in a greater self-selection bias in the 2007 survey compared to the 1997 survey. Likewise, there were several differences in survey administration and assessment of DSM-IV criteria including differences in question content that could potentially influence differences between the two surveys. However, the influence of these differences at the latent internalizing level are likely to be minor due to acceptable levels of measurement invariance as well as the use of the K10 items as factor indicators to model the latent internalizing construct. The K10 was measured in a consistent manner across both survey administrations and served as a sufficient anchor to link both surveys.

A final limitation is the inability of the current study to focus on any mortality effects that may be driving the change in psychopathology. For instance, it is a common finding that higher rates of psychopathology are related to higher rates of mortality, physical disorder, and poor health outcomes (Angst, Stassen, Clayton, & Angst, 2002; Kouzis, Eaton, & Leaf, 1995), suggesting that the minor decreases in psychopathology as the population ages might be due to the surviving population having lower levels of psychopathology at the outset of the study period. Evidence from longitudinal studies would be required to assess the role of selective mortality.

Several conclusions regarding age differences in internalizing psychopathology can be tentatively drawn based on the findings from this study. First, there are some differences across the ten year time period amongst individuals from the same birth cohort in the internalizing construct but these differences are relatively minor and are not reflective of systematic differences in responses to survey questions. Instead, the majority of individuals in the general population will experience constant levels of internalizing psychopathology, suggesting that the construct is relatively stable, at least across a 10-year time period. That being said, minor differences between age groups exist in the population suggesting that common factors or life events, which affect the population at various ages, can have the potential to alter the severity level of internalizing psychopathology to a smaller degree. The specific factors that could bring about these changes in the transitions to middle age and older age require further investigation.

## REFERENCES

Allemand, M., Zimprich, D., & Hendriks, A. A. (2008). Age differences in five personality domains across the life span. *Developmental Psychology, 44*(3), 758-770. doi: 10.1037/0012-1649.44.3.758

Andrews, G., Henderson, S., & Hall, W. (2001). Prevalence, comorbidity, disability and service utilisation: Overview of the Australian National Mental Health Survey. *British Journal of Psychiatry, 178*(2), 145-153. doi: 10.1192/bjp.178.2.145

Andrews, G., & Peters, L. (1998). The psychometric properties of the Composite International Diagnostic Interview. *Social Psychiatry and Psychiatric Epidemiology, 33*(2), 80-88.

Andrews, G., & Slade, T. (2001). Interpreting scores on the Kessler Psychological Distress Scale (K10). *Australian and New Zealand Journal of Public Health, 25*, 494-497.

Angst, F., Stassen, H. H., Clayton, P. J., & Angst, J. (2002). Mortality of patients with mood disorders: follow-up over 34-38 years. *Journal of Affective Disorders, 68*(2-3), 167-181.

Arnaud, B., Malet, L., Teissedre, F., Izaute, M., Moustafa, F., Geneste, J., . . . Brousse, G. (2010). Validity study of Kessler's Psychological Distress Scales conducted among patients admitted to French emergency department for alcohol consumption-related disorders. *Alcoholism: Clinical and experimental research, 34*(7), 1235-1245.

Baillie, A. (2005). Predictive gender and education bias in Kessler's psychological distress scale (K10). *Social Psychiatry and Psychiatric Epidemiology, 40*, 743-748.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Brault, M-C., Meuleman, B., & Bracke, P. (2012). Depressive symptoms in the Belgian population: disentangling age and cohort effects. *Social Psychiatry and Psychiatric Epidemiology, 47*(6), 903-915.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models*. Beverly Hills, CA: Sage.

Burns, R. A., Butterworth, P., Luszcz, M. A., & Anstey, K. J. (2013). Stability and change in level of probable depression and depressive symptoms in a sample of middle and older-aged adults. *International Psychogeriatrics, 25*(2), 303-309.

Cheung, G., & Rensvold, R. (1999). Testing factorial invariance across groups: a reconceptualization and proposed new method. *Journal of Management, 25*, 1-27.

Donker, T., Comijs, H., Cuijpers, P., Terluin, B., Nolen, W., Zitman, F., & Penninx, B. (2010). The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. *Psychiatry Research, 176*(1), 45-50.

Eaton, N. R., Krueger, R. F., & Oltmanns, T. F. (2011). Aging and the structure and long-term stability of the internalizing spectrum of personality and psychopathology. *Psychology and Aging, 26*, 987-993.

Eaton, N. R., Kruger, R. F., Markon, K. E., Keyes, K. M., Skodol, A. E., Wall, M., Hasin, D. S., & Grant, B. F. (2013). The structure and predictive validity of the internalizing disorders. *Journal of Abnormal Psychology, 122*, 86-92.

Fu, T. S-T., Lee, C-H., Gunnell, D., Lee, W-C., & Cheng, A, T-A. (2013). Changing trends in the prevalence of common mental disorders in Taiwan: a 20-year repeated cross-sectional survey. *The Lancet, 381,* 235-241.

Furukawa, T. A., Kawakami, N., Saitoh, M., Ono, Y., Nakane, Y., Nakamura, Y., . . . Kikkawa, T. (2008). The performance of the Japanese version of the K6 and K10 in

the World Mental Health Survey Japan. *International Journal of Methods in Psychiatric Research, 17*(3), 152-158. doi: 10.1002/mpr.257

Griffith, J. W., Zinbarg, R. E., Craske, M. G., Mineka, S., Rose, R. D., Waters, A. M., & Sutton, J. M. (2010). Neuroticism as a common dimension in the internalizing disorders. *Psychological Medicine, 40*(7), 1125-1136.

Hides, L., Lubman, D. I., Devlin, H., Cotton, S., Aitken, C., Gibbie, T., & Hellard, M. (2007). Reliability and validity of the Kessler 10 and Patient Health Questionnaire among injecting drug users. *Australian and New Zealand Journal of Psychiatry, 41*(2), 166-168.

Hu, L. T., & Bentler, P. M. (1998). Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification. *Psychological Methods, 3*(4), 424-453.

Johnson, W., McGue, M., & Krueger, R. F. (2005). Personality stability in late adulthood: a behavioral genetic analysis. *Journal of Personality, 73*(2), 523-552.

Jokela, M., Batty, G. D., & Kivimaki, M. (2013). Ageing and the prevalence and treatment of mental health problems. *Psychological Medicine, Epub ahead of print* (Jan 16).

Jordanova, V., Wickramesinghe, C., Gerada, C., & Prince, M. (2004). Validation of two survey diagnostic interviews among primary care attendees: a comparison of CIS-R and CIDI with SCAN ICD-10 diagnostic categories. *Psychological Medicine, 34*(06), 1013-1024. doi: doi:10.1017/S0033291703001727

Jorm, A. F. (2000). Does old age reduce the risk of anxiety and depression? A review of epidemiological studies across the adult life span. *Psychological Medicine, 30*(01), 11-22.

Kessler, R. C., Abelson, J., Demler, O., Escobar, J. I., Gibbon, M., Guyer, M. E., . . . Zheng, H. (2004). Clinical calibration of DSM-IV diagnoses in the World Mental Health

(WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMH-CIDI). *International Journal of Methods in Psychiatric Research, 13*, 122-139.

Kessler, R. C., Birnbaum, H. G., Bromet, E., Hwang, I., Sampson, N., & Shahly, V. (2010a). Age differences in major depression: results from the National Comorbidity Survey Replication (NCS-R). *Psychological Medicine, 40*(2), 225-237.

Kessler, R. C., Birnbaum, H. G., Shahly, V., Bromet, E., Hwang, I., McLaughlin, K. A., . . . Stein, D. J. (2010b). Age differences in the prevalence and co-morbidity of DSM-IV major depressive episodes: Results from the WHO world mental health survey initiative. *Depression and Anxiety, 27*(4), 351-364.

Kessler, R. C., Green, J. G., Gruber, M. J., Sampson, N. A., Bromet, E., Cuitan, M., . . . Zaslavsky, A. M. (2010c). Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. *International Journal of Methods in Psychiatric Research, 19*(S1), 4-22. doi: 10.1002/mpr.310

Kessler, R. C., & Ustun, T. B. (2004). The World Mental Health (WMH) survey initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research, 13*, 93-121.

Knauper, B., & Wittchen, H. (1994). Diagnosing major depression in the elderly: evidence for response bias in standardised diagnostic interviews? *Journal of Psychiatric Research, 28*, 147-164.

Kouzis, A., Eaton, W. W., & Leaf, P. J. (1995). Psychopathology and mortality in the general population. *Social Psychiatry and Psychiatric Epidemiology, 30*(4), 165-170.

Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry, 56*, 921-926.

Krueger, R. F., Caspi, A., Moffitt, T. E., & Silva, P. A. (1998). The structure and stability of common mental disorders (DSM-III-R): A longitudinal-epidemiological study. *Journal of Abnormal Psychology*, *107*, 216-227.

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: a model-based approach to understanding and classifying psychopathology. *Annual Reviews in Clinical Psychology, 2*, 111-133.

Lewinsohn, P. M., Rohde, P., Seeley, J. R., & Fischer, S. A. (1993). Age-cohort changes in the lifetime occurrence of depression and other mental disorders. *Journal of Abnormal Psychology, 102*(1), 110-120. doi: 10.1037/0021-843x.102.1.110

Lucas, R. E. & Donnellan, M. B. (2011). Personality development across the life span: longitudinal analyses with a national sample from Germany. *Journal of Personality and Social Psychology, 101*(4), 847-861.

McEvoy, P. M., Grove, R., & Slade, T. (2011). Epidemiology of anxiety disorder in the Australian general population: findings of the 2007 Australian National Survey of Mental Health and Wellbeing. *Australian and New Zealand Journal of Psychiatry, 45*, 957-967.

Mroczek, D. K., & Spiro, A. (2003). Modeling intraindividual change in personality traits: findings from the normative aging study. *Journal of Gerontology: Psychological sciences, 58B*, 153-165.

Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: multiple-group and growth modeling in Mplus. *Muthén & Muthén. Mplus Web Notes no 4*.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus Users' Guide.* Los Angeles, CA.: Muthén & Muthén.

O'Connor, D. W., & Parslow, R. A. (2009). Different responses to K10 and CIDI suggest that complex structured psychiatric interviews underestimate rates of mental disorder in older people. *Psychological Medicine, 39,* 1527-1532.

O'Connor, D. W., & Parslow, R. A. (2010a). Differences in older people's responses to CIDI's depression screening and diagnostic questions may point to age-related bias. *Journal of Affective Disorders, 125,* 361-364.

O'Connor, D. W., & Parslow, R. A. (2010b). Mental health scales and psychiatric diagnoses: responses to GHQ-12, K-10 and CIDI across the lifespan. *Journal of Affective Disorders, 121,* 263-267.

Simpson, K. R. S., Meadows, G. N., Frances, A. J., & Patten, S. B. (2012). Is mental health in the Canadian population changing over time? *Canadian Journal of Psychiatry, 57*(5), 324-331.

Slade, T., Grove, R., & Burgess, P. (2011). Kessler Psychological Distress Scale: normative data from the 2007 Australian National Survey of Mental Health and Wellbeing. *Australian and New Zealand Journal of Psychiatry, 45*(4), 308-316. doi: doi:10.3109/00048674.2010.543653

Slade, T., Johnston, A., Oakley Browne, M. A., Andrews, G., & Whiteford, H. (2009). 2007 National Survey of Mental Health and Wellbeing: Methods and key findings. *Australian and New Zealand Journal of Psychiatry, 43*(7), 594-605.

Slade, T., & Watson, D. (2006). The structure of common DSM-IV and ICD-10 mental disorder in the Australian general population. *Psychological Medicine,* 1-8.

Smith, H. L. (2008). Advances in Age-Period-Cohort analysis. *Sociological methods and research, 36*(3), 287-296.

Snowdon, J. (2001). Is depression more prevalent in old age? *Australian and New Zealand Journal of Psychiatry, 35*, 782-787.

Spiers, N., Bebbington, P. E., McManus, S., Brugha, T. S., Jenkins, R., & Meltzer, H. (2011). Age and birth cohort differences in the prevalence of common mental disorder in England: National Psychiatric Morbidity Surveys 1993-2007. *British Journal of Psychiatry, 198*(6), 479-484.

Spies, G., Stein, D. J., Roos, A., Faure, S. C., Mostert, J., Seedat, S., & Vythilingum, B. (2009). Validity of the Kessler 10 (K-10) in detecting DSM-IV defined mood and anxiety disorders among pregnant women. *Archives of Women's Mental Health, 21*, 1163-1168.

Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.

Sunderland, M., Hobbs, M. J., Anderson, T. M., & Andrews, G. (2012). Psychological distress across the lifespan: examining age-related bias in the Kessler-6 Psychological Distress Scale. *International Psychogeriatrics, 24*, 231-242.

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 Behavioural Assessment, 34*, 253-259.

Sunderland, M., Slade, T., Stewart, G., & Andrews, G. (2011). Estimating the prevalence of DSM-IV mental illness in the Australian general population using the Kessler Psychological Distress scale. *Australian and New Zealand Journal of Psychiatry, 45*, 880-889.

Thomas, K. L. (1998). The contribution of psychiatric epidemiology to planning better health care: lessons from Scandinavia. *Australasian Psychiatry, 6*(4), 178-180.

Trollor, J. N., Anderson, T. M., Sachdev, P. S., Brodaty, H., & Andrews, G. (2007). Age shall not weary them: Mental health in the middle-aged and the elderly. *Australian and New Zealand Journal of Psychiatry*, *41*(7), 581-589.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4-70.

Vollebergh, W. A. M., Iedema, J., Bijl, R. V., deGraaf, R., Smit, F., & Ormel, J. (2001). The structure and stability of common mental disorders. *Archives of General Psychiatry, 58*, 597-603.

World Health Organization. (1997). *CIDI-Auto 2.1: Administrator's guide and reference*. Sydney: World Health Organization.

**Table 1: Confirmatory factor analysis fit indices for the five tested models.**

Model fit in the 1997 survey

| Model | CFI | RMSEA | RMSEA 90% CI |
|---|---|---|---|
| Model A | 0.931 | 0.072 | 0.071-0.074 |
| **Model B** | **0.985** | **0.034** | **0.032-0.036** |
| Model C | 0.931 | 0.072 | 0.071-0.074 |
| Model D | 0.986 | 0.033 | 0.031-0.035 |
| Model E | 0.714 | 0.149 | 0.147-0.150 |

Model fit in the 2007 survey

| Model | CFI | RMSEA | RMSEA 90% CI |
|---|---|---|---|
| Model A | 0.942 | 0.055 | 0.053-0.057 |
| **Model B** | **0.992** | **0.021** | **0.019-0.023** |
| Model C | 0.942 | 0.055 | 0.053-0.057 |
| Model D | 0.993 | 0.020 | 0.018-0.022 |
| Model E | 0.864 | 0.085 | 0.083-0.088 |

Model fit in the merged sample

| Model | CFI | RMSEA | RMSEA 90% CI |
|---|---|---|---|
| Model A | 0.939 | 0.057 | 0.055-0.058 |
| **Model B** | **0.987** | **0.026** | **0.025-0.028** |
| Model C | 0.940 | 0.057 | 0.055-0.058 |
| Model D | 0.988 | 0.025 | 0.024-0.027 |
| Model E | 0.790 | 0.107 | 0.106-0.109 |

Notes: CFI = comparative fit index, RMSEA = root mean square error of approximation, RMSEA 90% CI = 90% confidence intervals surrounding the root mean square error of approximation. Model A = one-factor “internalizing” model, Model B = one-factor “internalizing” model with correlated errors between K10 items “depressed” and “so depressed”, “restless” and “so restless”, and “nervous” and “so nervous”, Model C = two-factor “Disorder – Distress” model, Model D = two-factor “Disorder – Distress” model with correlated errors between K10 items, and Model E = a two-factor “Disorder – Distress” model with factor covariance fixed to 0.

**Table 2: Multi-group CFA model fit statistic comparing measurement invariance models (based on model B) across survey administrations**

| Model | # | CFI | ΔCFI | ΔCFI-CI | RMSEA | 90% CI |
|---|---|---|---|---|---|---|
| Configural invariance | 126 | 0.987 | - | - | 0.030 | 0.028-0.031 |
| Scalar invariance | 83 | 0.987 | 0.000 | 0.000 | 0.027 | 0.025-0.028 |
| Strict invariance | 68 | 0.986 | -0.001 | -0.001 | 0.027 | 0.026-0.028 |
| Strict invariance with fixed residual covariances | 65 | 0.986 | 0.000 | -0.001 | 0.027 | 0.026-0.028 |

Notes: # = number of free parameters, CFI = comparative fit index, ΔCFI = difference in CFI values between successive models, ΔCFI-CI = difference in CFI values between configural invariance model, RMSEA = root mean square error of approximation, 90% CI = 90% confidence interval of RMSEA.

**Table 3: Multi-group CFA model fit statistic comparing measurement invariance models (based on model B) across ten year age bands**

| Model | # | CFI | ΔCFI | ΔCFI-CI | RMSEA | 90% CI |
|---|---|---|---|---|---|---|
| Configural invariance | 630 | 0.989 | - | - | 0.029 | 0.028-0.031 |
| Scalar invariance | 243 | 0.988 | -0.001 | -0.001 | 0.026 | 0.024-0.027 |
| Strict invariance | 108 | 0.985 | -0.003 | -0.004 | 0.027 | 0.026-0.028 |
| Strict invariance with fixed residual covariances | 81 | 0.985 | 0.000 | -0.004 | 0.027 | 0.026-0.028 |
| Strict invariance with fixed residual covariances and latent variances | 72 | 0.984 | -0.001 | -0.005 | 0.028 | 0.027-0.029 |
| Strict invariance with fixed residual covariances, latent variances, and latent means | 63 | 0.983 | -0.001 | -0.006 | 0.028 | 0.027-0.030 |

Notes: # = number of free parameters, CFI = comparative fit index, ΔCFI = difference in CFI values between successive models, ΔCFI-CI = difference in CFI values between configural invariance model, RMSEA = root mean square error of approximation, 90% CI = 90% confidence interval of RMSEA.

**Table 4: Mean deviation scores in latent internalizing between the same cohorts measured ten years apart (Model B).**

| Birth cohort | Age in 1997 | Age in 2007 | *d* | *p*-value |
|---|---|---|---|---|
| 1968-1977 | 20-29 | 30-39 | 0.05 | 0.28 |
| 1958-1967 | 30-39 | 40-49 | 0.13 | 0.03 |
| 1948-1957 | 40-49 | 50-59 | -0.03 | 0.65 |
| 1938-1947 | 50-59 | 60-69 | -0.18 | 0.01 |

*d* = mean difference between latent internalizing psychopathology across birth cohorts, *p*-value = significance of mean latent difference.

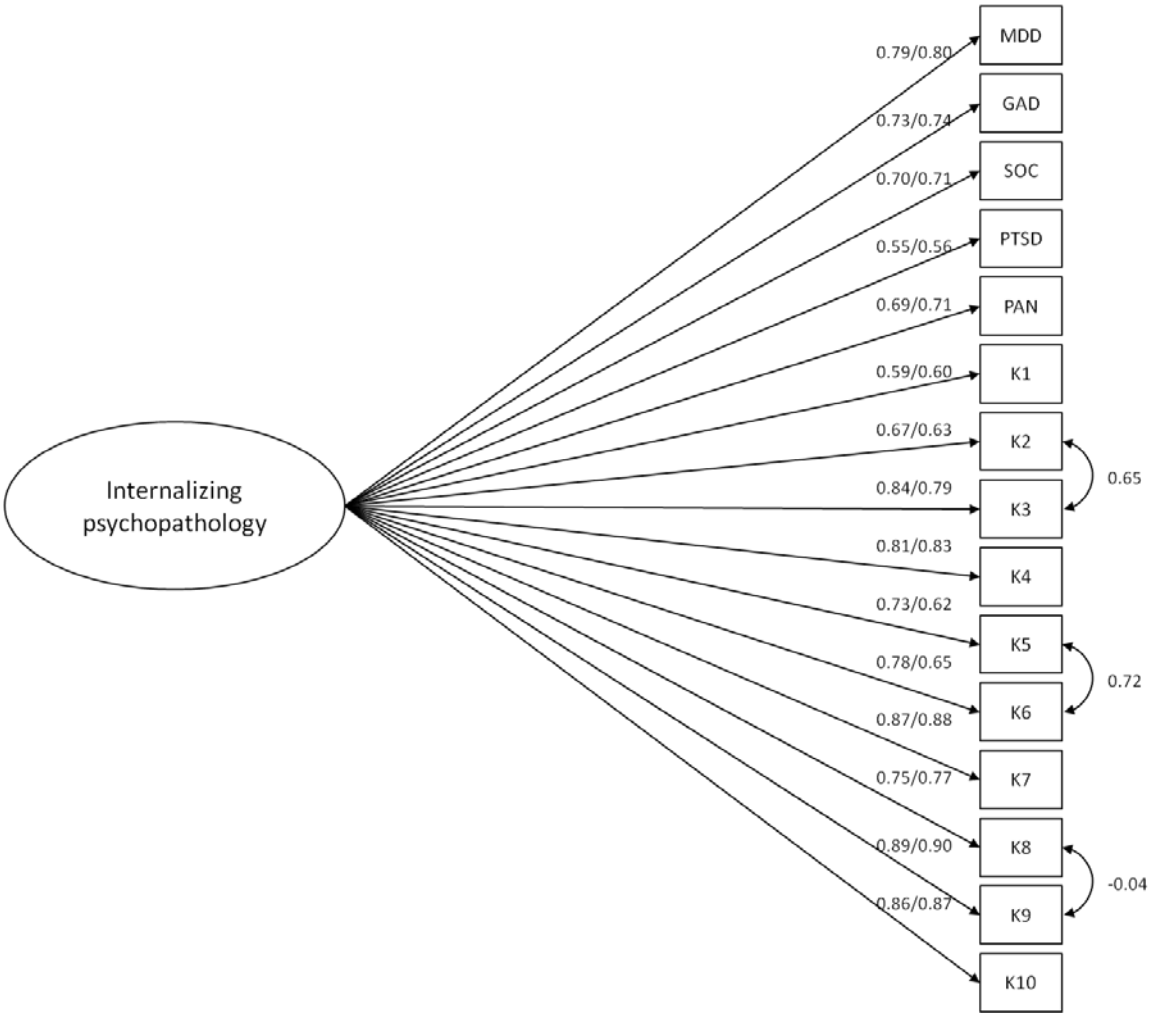


Figure 1: Best fitting EFA and CFA factor model (Model B) and loadings in merged sample. Loadings on the left represent the EFA model, loadings on the right represent the CFA model. Correlated residual errors were only specified in the CFA model. MDD = major depressive disorder, GAD = generalised anxiety disorder, SOC = social phobia, PTSD = post-traumatic stress disorder, PAN = panic disorder with/without agoraphobia, K1 – K10 = items one through ten on the Kessler psychological distress scale.

## ONLINE SUPPLEMENTARY MATERIAL

**eTable 1: Sample size and response rates for each factor indicator by survey administration and age group**

| | | 1997 Survey | | | | | 2007 Survey | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Age group** | | **20-29** | **30-39** | **40-49** | **50-59** | **60-69** | **20-29** | **30-39** | **40-49** | **50-59** | **60-69** |
| N | | 1849 | 2452 | 2153 | 1535 | 1170 | 1325 | 1618 | 1382 | 1294 | 1246 |
| MDD (%) | | 3.0 | 3.5 | 3.7 | 4.7 | 2.5 | 1.4 | 1.8 | 3.9 | 1.9 | 1.8 |
| SOC (%) | | 1.5 | 1.6 | 2.1 | 1.7 | 0.6 | 3.0 | 2.5 | 3.3 | 2.7 | 1.2 |
| GAD (%) | | 2.3 | 3.4 | 3.6 | 3.2 | 1.7 | 0.6 | 1.1 | 0.8 | 1.6 | 0.7 |
| PTSD (%) | | 1.4 | 0.8 | 0.9 | 1.1 | 0.2 | 2.3 | 3.3 | 3.4 | 2.4 | 1.7 |
| PAN (%) | | 1.3 | 1.9 | 2.3 | 1.7 | 0.9 | 1.0 | 1.3 | 1.4 | 0.9 | 1.1 |
| K1 | None of the time (%) | 54.1 | 56.9 | 55.9 | 59.8 | 63.0 | 45.2 | 48.7 | 46.7 | 51.6 | 56.5 |
| | A little of the time (%) | 27.7 | 22.0 | 23.5 | 19.9 | 17.2 | 28.0 | 25.3 | 23.4 | 17.9 | 19.1 |
| | Some of the time (%) | 12.9 | 16.2 | 15.3 | 13.5 | 13.9 | 20.8 | 18.3 | 20.6 | 23.6 | 16.6 |
| | Most of the time (%) | 4.5 | 4.0 | 4.1 | 5.1 | 4.4 | 4.7 | 6.0 | 7.3 | 5.1 | 4.8 |
| | All of the time (%) | 0.8 | 0.9 | 1.3 | 1.8 | 1.5 | 1.4 | 1.6 | 2.0 | 1.8 | 3.0 |
| K2 | None of the time (%) | 57.7 | 60.8 | 60.2 | 65.4 | 70.6 | 46.5 | 50.8 | 55.7 | 59.7 | 72.6 |
| | A little of the time (%) | 31.4 | 28.3 | 27.3 | 21.2 | 18.9 | 37.8 | 29.5 | 28.5 | 25.4 | 16.0 |
| | Some of the time (%) | 7.9 | 8.5 | 9.6 | 10.0 | 8.2 | 11.6 | 16.3 | 12.0 | 9.6 | 7.3 |
| | Most of the time (%) | 2.3 | 1.8 | 2.1 | 2.7 | 1.6 | 3.8 | 2.9 | 2.6 | 4.2 | 2.1 |
| | All of the time (%) | 0.7 | 0.7 | 0.8 | 0.6 | 0.7 | 0.3 | 0.5 | 1.2 | 1.1 | 2.0 |
| K3 | None of the time (%) | 92.0 | 93.3 | 93.1 | 93.2 | 95.6 | 90.7 | 91.7 | 92.0 | 94.0 | 95.2 |
| | A little of the time (%) | 5.8 | 4.8 | 4.2 | 4.1 | 3.3 | 6.6 | 5.4 | 4.7 | 3.5 | 2.9 |
| | Some of the time (%) | 1.6 | 1.4 | 1.7 | 1.5 | 0.8 | 1.9 | 2.2 | 2.1 | 1.6 | 1.5 |
| | Most of the time (%) | 0.6 | 0.5 | 0.6 | 1.0 | 0.1 | 0.5 | 0.4 | 1.1 | 0.7 | 0.4 |
| | All of the time (%) | 0.1 | 0.2 | 0.4 | 0.2 | 0.1 | 0.3 | 0.2 | 0.1 | 0.2 | 0.1 |
| K4 | None of the time (%) | 80.9 | 80.1 | 83.7 | 84.8 | 89.0 | 76.8 | 78.7 | 78.5 | 84.7 | 86.1 |
| | A little of the time (%) | 13.3 | 13.4 | 9.4 | 7.9 | 6.3 | 15.8 | 13.6 | 11.7 | 8.7 | 8.0 |
| | Some of the time (%) | 4.4 | 4.4 | 4.9 | 5.0 | 2.6 | 5.1 | 6.0 | 6.6 | 4.1 | 3.8 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Most of the time (%) | 1.1 | 1.5 | 1.5 | 1.3 | 1.3 | 1.8 | 1.5 | 2.1 | 1.7 | 0.9 |
| | All of the time (%) | 0.3 | 0.6 | 0.6 | 1.1 | 0.8 | 0.5 | 0.3 | 1.2 | 0.9 | 1.4 |
| K5 | None of the time (%) | 43.0 | 47.5 | 52.2 | 56.4 | 64.4 | 42.1 | 51.1 | 55.1 | 58.8 | 66.5 |
| | A little of the time (%) | 34.5 | 30.8 | 28.9 | 25.2 | 21.9 | 29.5 | 27.9 | 23.8 | 20.3 | 19.2 |
| | Some of the time (%) | 17.6 | 16.6 | 14.2 | 14.0 | 10.5 | 19.3 | 14.6 | 13.6 | 15.9 | 10.7 |
| | Most of the time (%) | 3.9 | 4.2 | 3.8 | 3.2 | 2.5 | 6.4 | 4.5 | 5.3 | 3.2 | 2.7 |
| | All of the time (%) | 0.9 | 0.8 | 0.9 | 1.2 | 0.8 | 2.7 | 1.8 | 2.1 | 1.8 | 0.9 |
| K6 | None of the time (%) | 78.2 | 79.0 | 83.0 | 84.4 | 89.8 | 74.6 | 79.3 | 80.5 | 85.1 | 87.3 |
| | A little of the time (%) | 13.5 | 12.4 | 10.1 | 8.2 | 5.0 | 14.8 | 11.5 | 8.7 | 6.7 | 6.2 |
| | Some of the time (%) | 6.3 | 6.3 | 4.3 | 4.4 | 3.5 | 7.5 | 6.4 | 7.5 | 5.8 | 4.2 |
| | Most of the time (%) | 1.6 | 1.7 | 1.9 | 2.6 | 1.5 | 2.1 | 2.2 | 2.2 | 1.7 | 2.0 |
| | All of the time (%) | 0.5 | 0.7 | 0.7 | 0.5 | 0.2 | 1.0 | 0.6 | 1.1 | 0.8 | 0.3 |
| K7 | None of the time (%) | 57.4 | 59.1 | 60.1 | 65.9 | 69.8 | 64.6 | 65.3 | 63.3 | 69.2 | 75.1 |
| | A little of the time (%) | 28.8 | 26.8 | 25.1 | 21.4 | 20.3 | 23.4 | 21.1 | 18.2 | 18.6 | 14.0 |
| | Some of the time (%) | 10.1 | 10.9 | 10.8 | 8.6 | 7.2 | 9.3 | 10.4 | 13.3 | 8.8 | 7.5 |
| | Most of the time (%) | 3.2 | 2.5 | 3.2 | 2.8 | 2.1 | 1.6 | 2.4 | 4.1 | 2.6 | 2.5 |
| | All of the time (%) | 0.5 | 0.8 | 0.8 | 1.2 | 0.6 | 1.1 | 0.8 | 1.1 | 0.9 | 0.9 |
| K8 | None of the time (%) | 57.3 | 56.0 | 59.8 | 63.3 | 68.4 | 56.5 | 55.5 | 55.1 | 60.7 | 72.7 |
| | A little of the time (%) | 27.6 | 27.7 | 23.8 | 20.7 | 16.7 | 27.6 | 26.4 | 24.3 | 20.7 | 14.2 |
| | Some of the time (%) | 10.0 | 11.3 | 10.1 | 9.6 | 9.9 | 11.9 | 12.8 | 13.7 | 13.7 | 8.2 |
| | Most of the time (%) | 4.6 | 3.7 | 4.9 | 4.5 | 3.2 | 2.6 | 3.4 | 4.8 | 3.6 | 3.0 |
| | All of the time (%) | 0.9 | 1.4 | 1.4 | 1.9 | 1.8 | 1.4 | 1.9 | 2.1 | 1.3 | 1.8 |
| K9 | None of the time (%) | 83.2 | 83.2 | 83.4 | 86.1 | 88.2 | 88.4 | 87.7 | 86.2 | 88.7 | 92.3 |
| | A little of the time (%) | 11.8 | 10.8 | 10.0 | 7.8 | 6.9 | 7.2 | 7.4 | 6.9 | 6.2 | 3.4 |
| | Some of the time (%) | 4.0 | 4.3 | 4.7 | 3.9 | 3.1 | 2.7 | 3.2 | 5.0 | 3.4 | 2.7 |
| | Most of the time (%) | 0.7 | 1.3 | 1.6 | 1.6 | 1.4 | 1.3 | 1.6 | 1.4 | 1.4 | 1.2 |
| | All of the time (%) | 0.3 | 0.4 | 0.4 | 0.6 | 0.4 | 0.4 | 0.0 | 0.5 | 0.4 | 0.3 |
| K10 | None of the time (%) | 87.8 | 87.2 | 87.4 | 88.9 | 90.5 | 84.7 | 87.3 | 82.6 | 87.3 | 89.0 |
| | A little of the time (%) | 8.4 | 8.2 | 7.3 | 6.1 | 5.5 | 9.4 | 8.0 | 8.6 | 6.6 | 6.2 |
| | Some of the time (%) | 2.2 | 3.0 | 3.2 | 2.6 | 2.0 | 4.0 | 3.2 | 5.6 | 4.3 | 3.0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Most of the time (%) | 1.4 | 1.2 | 1.5 | 1.6 | 1.4 | 1.2 | 1.2 | 2.1 | 1.1 | 1.5 |
| All of the time (%) | 0.3 | 0.4 | 0.7 | 0.8 | 0.7 | 0.7 | 0.4 | 1.2 | 0.7 | 0.3 |

Notes: MDD = major depressive disorder, GAD = generalised anxiety disorder, SOC = social phobia, PTSD = post-traumatic stress disorder, PAN = panic disorder with/without agoraphobia, K1 – K10 = items one through ten on the Kessler psychological distress scale.