The psychometric properties of the Kessler Psychological Distress Scale (K6) in a general population sample of adolescents

Louise Mewton

University of New South Wales

Ronald C. Kessler

Harvard Medical School

Tim Slade

University of New South Wales

Megan J. Hobbs

University of New South Wales

Louise Brownhill

University of New South Wales

Louise Birrell

University of New South Wales

Zoe Tonks

University of New South Wales

Maree Teesson

University of New South Wales

Nicola Newton

University of New South Wales

Cath Chapman

University of New South Wales

Steve Allsop

Curtin University

Leanne Hides

Queensland Univesity of Technology

Nyanda McBride

Curtin University

Gavin Andrews

University of New South Wales

Louise Mewton, Clinical Research Unit for Anxiety and Depression, St Vincent's Hospital, University of New South Wales, Sydney, Australia; Ronald C. Kessler, Department of Health Care Policy, Harvard Medical School, Boston, USA; Tim Slade, NHMRC Centre for Research Excellence in Mental Health and Substance Use, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia; Megan J. Hobbs, Clinical Research Unit for Anxiety and Depression, St Vincent's Hospital, University of New South Wales, Sydney, Australia; Louise Brownhill, Clinical Research Unit for Anxiety and Depression, St Vincent's Hospital, University of New South Wales, Sydney, Australia; Louise Birrell, NHMRC Centre for Research Excellence in Mental Health and Substance Use, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia; Zoe Tonks, NHMRC Centre for Research Excellence in Mental Health and Substance Use, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia; Maree Teesson, NHMRC Centre for Research Excellence in Mental Health and Substance Use, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia; Nicola Newton, NHMRC Centre for Research Excellence in Mental Health and Substance Use, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia; Cath Chapman, NHMRC Centre for Research Excellence in Mental Health and Substance Use, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia; Steve Allsop, National Drug Research Institute, Curtin University, Perth, Australia; Leanne Hides, Queensland

University of Technology, Brisbane, Australia; Nyanda McBride, National Drug Research Institute, Curtin University, Perth, Australia; Gavin Andrews, Clinical Research Unit for Anxiety and Depression, St Vincent's Hospital, University of New South Wales, Sydney, Australia.

Corresponding concerning this article should be addressed to Louise Mewton, Clinical Research Unit for Anxiety and Depression, St Vincent's Hospital, University of New South Wales, Sydney, Australia. E-mail: louisem@unsw.edu.au

Abstract

The 6-item Kessler Psychological Distress Scale (K6) is a screener for psychological distress that has robust psychometric properties among adults. Given that a significant proportion of adolescents experience mental illness, there is a need for measures that accurately and reliably screen for mental disorders in this age group. This study examined the psychometric properties of the K6 in a large general population sample of adolescents (n = 4,434; mean age = 13.5 years; 44.6% male). Factor analyses were conducted to examine the dimensionality of the K6 in adolescents, and to investigate sex-based measurement invariance. This study also evaluated the K6 as a predictor of scores on the Strengths and Difficulties Questionnaire (SDQ). The K6 demonstrated high levels of internal consistency, with the six items loading primarily on one factor. Consistent with previous research, females reported higher mean levels of psychological distress when compared with males. The identification of sex-based measurement non-invariance in the item thresholds indicated that these mean differences most likely represented reporting bias in the K6 items, rather than true differences in the underlying psychological distress construct. The K6 was a fair to good predictor of abnormal scores on the SDQ, but predictive utility was relatively low amongst males. Future research needs to focus on refining and augmenting the K6 scale to maximize its utility in adolescents.

Introduction

4

The 6-item Kessler Psychological Distress Scale (K6) is a dimensional measure of non-specific psychological distress that has been used extensively in community epidemiological surveys of adults (Kessler et al., 2002). When compared with other screening scales, including the Composite International Diagnostic Interview (CIDI) - Short Form and the General Health Questionnaire 12-item, the K6 has superior sensitivity and specificity in terms of identifying cases of severe mental illness in adults (Furukawa, Kessler, Slade, & Andrews, 2003; Kessler et al., 2003). Consequently, the K6 has been translated into 14 languages, included in the World Health Organization World Mental Health Surveys (Kessler et al., 2002), and administered as part of recurring national household surveys in the United States, Australia and Canada (Sunderland, Slade, Stewart, & Andrews, 2011).

Although the psychometric properties of the K6 in adult populations are robust, there have been limited investigations of the reliability and validity of the K6 among adolescents. Three previous studies have investigated the distribution and internal consistency of the K6 in adolescents (Chan & Fung, 2014; Green, Gruber, Sampson, Zaslavsky, & Kessler, 2010; Peiper, Clayton, Wilson, & Illback, 2015). Similar to adults, the K6 has a J-shaped distribution among adolescents, with the majority of respondents reporting minimal psychological distress. Internal consistency of the K6 ranged from good to excellent (α = 0.78-0.90), with the K6 items assessing a single underlying psychological distress trait, which is consistent with adult data (Kessler et al., 2002). Despite the use of the K6 in large international studies of adolescents (Chan & Fung, 2014; Green et al., 2010; Huang, Xia, Sun, Zhang, & Wu, 2009; Newton, Andrews, Champion, & Teesson, 2014), more comprehensive psychometric investigations of this scale have not been conducted in this age group. In particular, the extent to which the K6 demonstrates measurement invariance with regards to key demographic variables has not been investigated in adolescents. Previous

studies of both adolescents (Li, Green, Kessler, & Zaslavsky, 2010) and adults (Drapeau et al., 2010; Jorm et al., 2005), have indicated that females have higher levels of psychological distress than males. The examination of measurement invariance can provide an indication of whether this disparity reflects true mean differences in the psychological distress construct, or bias in the way this construct is assessed. Sex-based measurement invariance of the K6 has been examined in an adult general population sample (Drapeau et al., 2010). Whilst this study supported sex-based measurement invariance overall, there was some indication of sex-based bias in the K6 items when applied to a sub-sample of younger adults (aged 18-39 years). These previous findings indicate that the K6 items may be problematic when applied to younger populations. The first aim of the current study was to therefore investigate sex-based measurement invariance of the K6 items in an adolescent sample.

The second aim of this study was to conduct a more nuanced investigation of the degree to which each of the K6 items, and the overall K6 scale, provided information about the psychological distress construct at varying levels of severity. To date, the reliability of the K6 in adolescents has been expressed in terms of a single index (i.e., Cronbach's alpha) under the assumption that the precision of this scale is constant along the entire underlying continuum of psychological distress. In the current study, modern psychometric methods informed by factor analysis were adopted to allow a more thorough investigation of the precision of the K6 items across the spectrum of psychological distress.

Given that one in every four to five adolescents experiences at least one mental disorder in any given year (Patel, Flisher, Hetrick, & McGorry, 2007) , and that the peak disability related to psychiatric disorders occurs in adolescence (Murray et al., 2013), reliable and valid screening scales for adolescent mental illness are needed as a means of accurately identifying

respondents in need of further evaluation. To this end, recent studies have focused on the use of the K6 as such a screening tool, investigating its utility in predicting affective, anxiety and behavioral disorders in adolescents. The K6 has been shown to be a fair predictor of affective [area under the receiver operating curve (AUC) = 0.77] and anxiety (AUC = 0.73) disorders, as diagnosed using a modified version of the CIDI, but a poor predictor of behavior disorders (AUC = 0.67) (Green et al., 2010). The K6 has also been found to be an excellent predictor of major depression as diagnosed by the Beck Depression Inventory II (AUC = 0.90) (Chan & Fung, 2014). Recent research has also focused on the utility of the K6 in predicting serious emotional disturbance (SED) in adolescents (Green et al., 2010; Li et al., 2010; Linden, Phillips, & Leclerc, 2007). In these studies, SED has been defined as the presence of one or more mental disorders in addition to either: 1) significant disorder-related impairment; 2) the presence of bipolar I disorders; or 3) a suicide attempt in the previous 12 months. The K6 was found to be a fair predictor of SED (AUC = 0.74) (Green et al., 2010), whilst the correlation between the K6 and SED appears to be moderate ($\rho = 0.52$) (Li et al., 2010). In an extension of this previous work, the current study evaluated the K6 as a predictor of scores on the Strengths and Difficulties Questionnaire (SDQ), one of the most frequently used screening instruments for child and adolescent mental health problems worldwide (Goodman, 1997; Goodman, Ford, Simmons, Gatward, & Meltzer, 2000; Goodman, Meltzer, & Bailey, 2003).

In order to address these objectives, we analyzed data from 4,434 adolescents participating in the [redacted for review] study, a large cluster-randomized controlled trial of the efficacy of an online, school-based universal prevention program for mental health and substance use problems (Teesson et al., 2014). The current study focused on descriptive data collected at baseline, prior to the commencement of the intervention.

Methods

Design

A cluster-randomized controlled trial was implemented and convenience sampling was used to recruit 68 schools from [redacted for review]. Data were collected both online and in paper and pencil format. The current study reports on baseline data collected online from a subsample of schools (n=54). The sample consisted of students from state (n = 27), independent (n = 17) and catholic (n = 10) schools. Baseline data was collected from Year 8 students between January 2014 and May 2014. All aspects of this trial were approved by the relevant education departments and university Human Research Ethics Committees. The trial is registered with the Australian Clinical Trials Registry (ACTRN12613000723785).

Participants

For those students completing online questionnaires, informed consent was obtained from 4,773 adolescents and their parents. A total of 4,434 adolescents completed their baseline assessments, including K6 data. Limited demographic information was collected, including age, sex and country of birth. The mean age of the sample was 13.5 (0.56) years, 44.6% were male and 83.7% were born in Australia. There were no statistically significant differences between males and females in terms of age or country of birth.

Measures

For the purposes of this study, only the K6 and the SDQ, as described in detail below, were analyzed.

The K6 consists of six questions that ask respondents how frequently in the past 30 days they had felt: 1) nervous; 2) hopeless; 3) restless or fidgety (restless); 4) so depressed that nothing

could cheer them up (depressed); 5) that everything was an effort (effort); and 6) worthless. For each of these questions, the K6 included five response options: 'never', 'a little of the time', 'some of the time', 'most of the time' and 'all of the time'. Responses were scored in the range of 0 ('never') to 4 ('all of the time'), generating a scale with a range of 0–24 (Kessler et al., 2002).

The SDQ is a 25-item scale designed to screen for maladaptive and prosocial behaviors among children and adolescents. Although multi-informant reporting has been recommended for children aged younger than 11 years, the psychometric properties of the self-report version among adolescents aged 11-16 years have been widely established (Goodman, 1997; Goodman et al., 2000; Goodman et al., 2003). The SDQ has five subscales measuring emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and positive prosocial behaviors over the last six months using the response options 'not true', 'somewhat true' and 'certainly true'. Scores from the emotional symptoms, conduct problems, hyperactivity/inattention and peer relationship problems range from 0-2 and can be summed to generate a total difficulties score ranging from 0-40. Based on epidemiological data, difficulties scores have been provisionally banded into 'normal' (0-15), 'borderline' (16-19) and 'abnormal' (20-40). Each of the subscales has also been banded into 'normal', 'borderline' and 'abnormal' ranges. The self-report scale SDQ has been shown to reliably and validly detect behavioral and emotional problems in clinical and community samples of Australian adolescents (Hawes & Dadds, 2004; Mathai, Anderson, & Bourne, 2002).

Statistical analysis

Prevalence and distribution of the K6

Using SPSS version 21, the prevalence and cumulative prevalence of K6 scores (ranging from 0-24) were calculated by sex. Mean K6 scores, associated standard deviations and percentiles of the K6 score distribution were also calculated by sex. The reliability of the K6 (Cronbach's alpha) was also calculated in the whole sample, as well as by sex.

Factor Analysis

Whilst the K6 has been shown to be unidimensional in adult samples, the dimensionality of the K6 has rarely been investigated in samples of adolescents. Exploratory factor analyses (EFA) with various rotations were therefore conducted using MPlus Version 7.1 (Muthen & Muthen, 2001) to determine whether the K6 items measured a single underlying construct in this sample of adolescents. Unidimensionality of the K6 scale was supported if EFA revealed a large first eigenvalue and a second eigenvalue less than 1.0 as per previous analyses of the K6 (Kessler et al., 2010; Peiper et al., 2015). Confirmatory factor analyses (CFA) were then conducted in MPlus Version 7.1 using weighted least squares means and variance adjusted estimation as recommended for categorical variables (Muthen & Muthen, 2001). Model fit was assessed using the Root Mean Square Error of Approximation (RMSEA; ≤ 0.08 was considered adequate, ≤ 0.05 was considered very good), the Comparative Fit Index (CFI; ≥ 0.95 considered very good) and the Tucker Lewis Index (TLI; ≥ 0.95 considered very good) (Hu & Bentler, 1999). Choice of model fit statistics was informed by previous studies that indicate the chi-square goodness of fit statistic is overly sensitive to minor differences in very large samples (Browne, MacCallum, Kim, Andersen, & Glaser, 2002).

Measurement invariance

Having established the best-fitting model, sex-based measurement invariance of the K6 items was examined within a CFA framework. The investigation of measurement invariance within

a CFA framework involves placing successive constraints on factor loadings and item thresholds and comparing the fit of these constrained models with models where parameters are freely estimated (Millsap & Yun-Tein, 2004). Factor loadings represent the strength of the relationship between each of the K6 items and the underlying psychological distress factor. Factor loadings indicate the amount of information each item provides about psychological distress. The item thresholds, on the other hand, indicate the point along the underlying latent factor at which an individual would be expected to transition from one response category to the next. Each of the K6 items, for example, has four thresholds. The first threshold indicates the point at which an individual can be expected to transition from 'never' to 'a little of the time', the second threshold indicates the expected point of transition from 'a little of the time' to 'some of the time', the third threshold indicates the expected point of transition from 'some of the time' to 'most of the time', whilst the fourth threshold indicates the expected point of transition from 'most of the time' to 'all of the time'. The thresholds are therefore indicative of the severity of each of the K6 items and their response categories.

The first step in testing measurement invariance was to begin with the best-fitting CFA model and freely estimate factor loadings and item thresholds in both males and females (the baseline model). The next step was to estimate a model in which factor loadings were constrained to be equal across sex (weak invariance) (Meredith, 1993). The fit of this model was compared with the baseline model using chi-square difference testing (the MPlus DIFFTEST option). Finally, both factor loadings and item thresholds were constrained to be equal across sex (strong invariance) (Meredith, 1993). The fit of this model was then compared with the weak invariance model, again using chi-square difference testing (the MPlus DIFFTEST option). If strong invariance is supported, then sex-based differences on

the K6 reflect true differences in the underlying psychological distress factor, rather than bias in the assessment instrument. If strong invariance is not supported, then something other than the underlying psychological distress factor is contributing to any sex-based differences in K6 scores.

Item and test information functions

Item and test information functions were also constructed as a means of visually representing any sex-based measurement non-invariance identified. These information functions were also used to demonstrate the amount of information provided by each of the K6 items across various levels of underlying psychological distress (ranging from -3 to +3 standard deviations). For each item, the amount of information provided reflected the relationship between that item and the underlying psychological distress factor (i.e., the factor loading derived from the CFA), whilst the positioning of the information function along the x-axis reflected the severity of that item (i.e., the item thresholds derived from the CFA).

Relationship between the K6 and SDQ

Using Medcalc Version 15.6 (Schoonjans, Zalata, Depuydt, & Comhaire, 1995), the area under the curve (AUC) was calculated as a measure of the extent to which the K6 predicted an abnormal total difficulties score on the SDQ, as well as an abnormal score on each of the four maladaptive behavior SDQ subscales. An AUC of 0.5 would indicate that the K6 is no better than chance at predicting an abnormal score, whilst an AUC of 1.0 would indicate that the K6 predicts an abnormal score perfectly (Zweig & Campbell, 1993). According to established criteria, AUCs are typically interpreted as excellent (0.90-1.00), good (0.80-0.89), fair (0.70-0.79) and poor (<0.70) (Cicchetti, 2001). In order to investigate the relationship between the K6 and SDQ scales, each of the SDQ scales, as well as the total SDQ difficulties

score, were dichotomized into a normal/borderline (non-cases) category and an abnormal (cases) category as per the scoring rules outlined above. The AUC was calculated using these dichotomized SDQ scores and the continuous K6 score (with scores on the K6 ranging from 0-24 as described above).

Results

Prevalence and distribution of the K6

The distribution of the K6 (Figure 1a, Figure 1b, Figure 2 and Table 1) was broadly similar to that seen in adult samples, as well as adolescent samples in the US and China (Chan & Fung, 2014; Green et al., 2010; Peiper et al., 2015). Figure 3 shows the distribution of each of the K6 items by sex. Mean psychological distress scores were higher for females (mean K6 score = 6.15) when compared with males [mean K6 score = 5.28; $t(4429) = 6.16$, $p < 0.01$; Cohen's $d = 0.19$]. Internal consistency of the K6 items was high in the overall sample (Cronbach's α = 0.84; 95% CI: 0.83-0.85), as well as for males (Cronbach's α = 0.83; 95% CI: 0.82-0.84) and females (Cronbach's α = 0.85; 95% CI: 0.84-0.86) considered separately.

Factor Analysis

The K6 items were all significantly correlated with each other, with correlations ranging from 0.43-0.83 (Table 2). Exploratory factor analysis identified a strong first factor (eigenvalue = 3.86) with little support for additional factors (eigenvalues ranging from 0.16-0.70). Table 3 includes the fit statistics for the EFA using the oblique GEOMIN rotation. Alternative rotations (PROMAX and VARIMAX) were also tested, but these did not change the results appreciably. In terms of CFI and TLI, the fit of the 1-factor and 2-factor EFA models using the oblique GEOMIN rotation were both very good. The RMSEA indicated poor fit for the 1-factor EFA model and adequate fit for the 2-factor EFA model. In order to investigate the

dimensionality of the K6 further, several CFA models were implemented and the model fit statistics compared (Table 3). Model 1 in Table 3 represents the 1-factor CFA with all six items loading on a single factor. The RMSEA for this model was poor (0.114). Following a recent study investigating the K6 in a large school sample (Peiper et al., 2015), as well as studies of the K6 in adult populations (Drapeau et al., 2010), post hoc modification indices for the 1-factor model were inspected and correlated residuals were included in the model (the residuals of the Nervous item were correlated with the Restless, Depressed and Worthless items, whilst the residuals of the Effort item were correlated with Restless item). The fit statistics for this model (Model 1a) were all very good (see Table 3). A 2-factor CFA model was also investigated (Model 2). Guided by the results of the EFA, this included one factor defined by the Nervous, Restless and Effort items and another factor defined by the Hopeless, Depressed and Worthless items. The fit for this model was adequate, with an RMSEA of 0.76 (see Table 3). Guided by the EFA eigenvalues, CFA model fit indices and previous research on the dimensionality of the K6 in adolescents and adults, Model 1a was selected as the best fitting model.

Measurement invariance

The first step in the investigation of measurement invariance therefore involved fitting Model 1a for males and females separately, with no equality constraints specified (see Table 3 for fit statistics of the baseline model). The baseline model was then compared with a model that constrained factor loadings to equality across sex (weak invariance). In this model, the correlated residuals associated with Model 1a were also constrained to be equal across sex. There was no statistically significant differences between the baseline model and the weak invariance model [$\chi^2(9, N = 4434) = 7.48, p = .59$], indicating that the factor loadings and pre-specified correlated residuals were invariant by sex (see Table 3 for fit statistics of the

14

weak invariance model). The weak invariance model was then compared with a model that constrained item thresholds, in addition to factor loadings and pre-specified item residuals, to equality across sex (strong invariance model). There were statistically significant differences between the weak and strong invariance models [χ^2(24, N = 4434) = 286.93, p < 0.001], indicating that the thresholds were not invariant by sex (see Table 3 for fit statistics of the strong invariance model). Based on model modification indices, several partial strong invariance models were then specified. These models successively relaxed the equality constraints on various item thresholds across sex. Using chi-squared difference testing these partial invariance models were each compared with the weak invariance model. All of these models elicited statistically significant reductions in model fit, with the results ultimately indicating a lack of measurement invariance in the thresholds for each of the K6 items. The standardized factor loadings and thresholds for the weak invariance model are presented in Table 4. Inspection of the item thresholds indicated that, in general, the K6 items were more severe for males after taking into account the underlying level of psychological distress. In other words, given similar levels of psychological distress, males were less likely to endorse each of the K6 items than their female counterparts.

Item and test information functions

Item information functions incorporating the sex-based measurement non-invariance for each of the K6 items have been included in Figures 4a-f. These functions were derived from the factor loadings and thresholds established in CFA Model 1a. Within each item, the form of each of the curves is similar across sex which reflects the sex-based invariance of the factor loadings (i.e., weak invariance). For each item, however, the male curve is displaced along the x-axis, indicating that at any given level of underlying psychological distress the amount of information provided by the K6 items differed by sex. This is a reflection of the non-

invariance of the item thresholds (i.e., strong measurement non-invariance). The test information function incorporating the sex-based measurement non-invariance has also been included in Figure 5. This figure shows the amount of information the K6 scale (incorporating all six items) provides at various levels of underlying psychological distress (ranging from -3 to +3 standard deviations). Again, the form of the curves is similar, with the male curve displaced along the x-axis, providing a visual representation of the sex-based threshold non-invariance of the K6 items when considered as a scale.

These information functions also show the degree to which each of the K6 items, and the overall K6 scale, provide information about the psychological distress construct at varying levels of severity. As can be seen, the Nervous, Restless and Effort items all provided a relatively low amount of information, regardless of sex, reflecting the lower factor loadings between these items and the psychological distress factor. The amount of information provided by the Nervous, Restless and Effort items across the psychological distress factor were also similar, indicating a degree of inter-item redundancy. Inspection of the test information function reveals that the K6 scale provided maximum information at the more severe end of the psychological distress continuum. For females, the maximum amount of information provided by the scale was about one standard deviation above the mean, whilst for males the maximum amount of information provided was about two standard deviations above the mean.

The relationship between the K6 and SDQ

The numbers of participants scoring in the abnormal range of the SDQ scales are listed in Table 5. As can be seen from the AUC values in Table 6, the K6 provided fair predictions of abnormal scores on each of the SDQ subscales, except for the Emotion subscale, where the

AUC was in the good range. When investigating the AUC values by sex, the K6 was a better predictor for each of the SDQ subscales in females when compared with males. The K6 was a relatively poor predictor of SDQ scores in males, particularly in terms of the Conduct and Peer Difficulties subscales.

Discussion

The current study investigated the psychometric performance of the K6 in a sample of Australian adolescents. The distribution of K6 scores was similar to that found in adult general population samples (Andrews & Slade, 2001; Furukawa et al., 2003; Kessler et al., 2002), and adolescent samples (Chan & Fung, 2014; Green et al., 2010; Peiper et al., 2015). Overall, levels of psychological distress were low, with female adolescents reporting higher mean levels of psychological distress than their male counterparts. The internal consistency of the K6 was high and all six items primarily loaded onto a single factor representative of psychological distress.

Despite these promising results, further psychometric analyses indicated some problems with the application of the K6 to this adolescent sample. Measurement non-invariance was identified in the thresholds for each of the K6 items. After controlling for the underlying psychological distress factor, males were less likely to endorse each of the K6 items when compared with their female counterparts. This indicates that sex-based differences in K6 scores reflect bias in the reporting of the K6 items, rather than true differences in psychological distress. Previous analyses of adult samples have similarly found evidence of sex-based non-invariance in young adults (Drapeau et al., 2010), indicating that the K6 may be particularly problematic when applied to younger populations. Future research may focus on the intersection of sex- and age-based measurement invariance to determine whether the

development of alternative, adolescent-specific measures of psychological distress is warranted. An augmented, adolescent-specific K6 has been proposed which includes questions relating to behavioral disorders (Green et al., 2010). It is possible that these questions may be more relevant to males, balancing out the measurement non-invariance at least at the level of the K6 scale. The inclusion of behavioral symptoms in the augmented K6 may also increase its ability to predict scores on the SDQ, particularly in male samples where predictive power was relatively poor.

When constructing the K6, the items were selected based on unidimensionality and minimal redundancy among adults (Kessler et al., 2002), two properties which were not necessarily supported in the current study. EFA found a strong first factor and little evidence of a meaningful second factor, with eigenvalues highly consistent with those found in US adolescents (Green et al., 2010; Peiper et al., 2015). Whilst CFA supported a one factor model in the current study, optimal fit of this model was only achieved when correlated residuals based on post hoc modification indices were included. Previous studies of the K6 in adults and adolescents have similarly relied on correlated residuals to achieve optimal fit of a 1-factor model (Drapeau et al., 2010; Peiper et al., 2015), perhaps reflecting the fact that the K6 items represent interrelated aspects of psychological distress. The correlated residuals of the K6 items may also be indicative of inter-item redundancy. EFA revealed some high inter-correlations among the items, particularly with regards to the Hopeless, Depressed and Worthless items, indicating a degree of item redundancy. These high inter-correlations were not evident in a previous study of adolescents in the US (Green et al., 2010), and appear to be specific to the current sample. The similar item information functions of the Effort, Restless and Nervous items also indicated that these items provide largely redundant information in terms of both factor loadings and item thresholds. When applied to adolescents in the general

population, there appears to be some degree of redundancy in the K6 items, indicating that alternative candidate items may better capture the underlying spectrum of psychological distress in adolescent samples. In addition, the information provided by the Effort, Restless and Nervous items were low, indicating a weak relationship between these items and the underlying dimension of psychological distress. This coupled with the similar range of information provided by these items suggests that alternative items may be developed for optimal use of the K6 in adolescents.

Limitations of this study need to be noted. The [redacted for review] sample cannot be assumed to be representative of adolescents in the Australian general population. The timeframe for assessment for the K6 (past 30 days) and SDQ (past six months) differed and likely resulted in an underestimate of the resulting AUCs. A further limitation of the study is the reliance on participant self-report measurement, introducing the possibility of over or under reporting by participants. Although an array of multi-informant measures would be preferable, research has shown that adolescent self-report is reliable and valid (Clark & Winters, 2002; Van De Looij, Petra, & De Wilde, 2008).

The current study found that the distribution of the K6 in Australian adolescents is similar to that found in previous research focusing on both adult and adolescent samples. Consistent with previous research, females reported higher mean levels of psychological distress when compared with males. The identification of measurement non-invariance indicated that these mean differences most likely represent reporting bias in the K6 items, rather than true differences in the underlying psychological distress construct. The inclusion of behavioral symptoms, as has been investigated in previous research (Green et al., 2010), may reduce the impact of sex-based measurement non-invariance at the level of the K6 scale, and increase

the predictive utility of the K6 especially amongst males. When compared with other brief screening scales for adolescent mental illness, distress and impairment (Levitt, Saka, Hunter Romanelli, & Hoagwood, 2007), the K6 has a number of desirable properties, including brevity, clinical relevance, ease of scoring and self-administration (Green et al., 2010). Future research needs to focus on refining and augmenting the K6 scale to maximize its utility in adolescents.

Acknowledgements

<center>**References**</center>

Andrews, G., & Slade, T. (2001). Interpreting scores on the Kessler Psychological Distress

    Scale (K10). *Australian and New Zealand Journal of Public Health, 25*(6), 494-497.

Browne, M., MacCallum, R., Kim, C., Andersen, B., & Glaser, R. (2002). When fit indices

    and residuals are incompatible. *Psychological Methods, 7*(4), 403.

Chan, S. M., & Fung, T. C. T. (2014). Reliability and validity of K10 and K6 in screening

    depressive symptoms in Hong Kong adolescents. *Vulnerable Children and Youth*

    Studies, 9(1), 75-85.

Cicchetti, D. V. (2001). The precision of reliability and validity estimates re-visited:

    distinguishing between clinical and statistical significance of sample size

    requirements. *Journal of Clinical and Experimental Neuropsychology, 23*(5), 695-

    700.

Clark, D. B., & Winters, K. C. (2002). Measuring risks and outcomes in substance use

    disorders prevention research. *Journal of consulting and clinical psychology, 70*(6),

    1207.

Drapeau, A., Beaulieu-Prévost, D., Marchand, A., Boyer, R., Préville, M., & Kairouz, S.

    (2010). A life-course and time perspective on the construct validity of psychological

    distress in women and men. Measurement invariance of the K6 across gender. *BMC*

    Medical Research Methodology, 10(1), 68.

Furukawa, T. A., Kessler, R. C., Slade, T., & Andrews, G. (2003). The performance of the

    K6 and K10 screening scales for psychological distress in the Australian National

    Survey of Mental Health and Well-Being. *Psychological Medicine, 33*(2), 357-362.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal*

    of child psychology and psychiatry, 38(5), 581-586.

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry, 177*(6), 534-539.

Goodman, R., Meltzer, H., & Bailey, V. (2003). The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. *International Review of Psychiatry, 15*(1-2), 173-177.

Green, J. G., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2010). Improving the K6 short scale to predict serious emotional disturbance in adolescents in the USA. *International journal of methods in psychiatric research, 19*(S1), 23-35.

Hawes, D. J., & Dadds, M. R. (2004). Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Australian and New Zealand Journal of Psychiatry, 38*(8), 644-651.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 26*, 1-55.

Huang, J. P., Xia, W., Sun, C. H., Zhang, H. Y., & Wu, L. J. (2009). Psychological distress and its correlates in Chinese adolescents. *Australian and New Zealand Journal of Psychiatry, 43*(7), 674-681.

Jorm, A., Windsor, T., Dear, K., Anstey, K., Christensen, H., & Rodgers, B. (2005). Age group differences in psychological distress: the role of psychosocial risk factors that vary with age. *Psychological Medicine, 35*(09), 1253-1263.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L. T., . . . Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*(6), 959-976.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., . . . Walters, E. E. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60*(2), 184-189.

Kessler, R. C., Green, J. G., Gruber, M. J., Sampson, N. A., Bromet, E., Cuitan, M., . . . Hu, C. Y. (2010). 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.

Levitt, J. M., Saka, N., Hunter Romanelli, L., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology, 45*(2), 163-191.

Li, F., Green, J. G., Kessler, R. C., & Zaslavsky, A. M. (2010). Estimating prevalence of serious emotional disturbance in schools using a brief screening scale. *International journal of methods in psychiatric research, 19*(S1), 88-98.

Linden, W., Phillips, M. J., & Leclerc, J. (2007). Psychological treatment of cardiac patients: a meta-analysis. *European heart journal, 28*(24), 2972-2984.

Mathai, J., Anderson, P., & Bourne, A. (2002). The Strengths and Difficulties Questionnaire (SDQ) as a screening measure prior to admission to a Child and Adolescent Mental Health Service (CAMHS). *Advances in Mental Health, 1*(3), 235-246.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39*(3), 479-515.

Murray, C. J., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., . . . Abdalla, S. (2013). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21

regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*(9859), 2197-2223.

Muthen, B. O., & Muthen, L. K. (2001). *Mplus. Statistical Analysis with Latent Variables. User's Guide (v2.01)*. Los Angeles, CA.: Muthen and Muthen.

Newton, N. C., Andrews, G., Champion, K. E., & Teesson, M. (2014). Universal Internet-based prevention for alcohol and cannabis use reduces truancy, psychological distress and moral disengagement: A cluster randomised controlled trial. *Preventive medicine, 65*, 109-115.

Patel, V., Flisher, A. J., Hetrick, S., & McGorry, P. (2007). Mental health of young people: a global public-health challenge. *The Lancet, 369*(9569), 1302-1313.

Peiper, N., Clayton, R., Wilson, R., & Illback, R. (2015). The performance of the K6 Scale in a large school sample. *Psychological Assessment, 27*(1), 228.

Schoonjans, F., Zalata, A., Depuydt, C., & Comhaire, F. (1995). MedCalc: a new computer program for medical statistics. *Computer methods and programs in biomedicine, 48*(3), 257-262.

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*(10), 880-889.

Teesson, M., Newton, N. C., Slade, T., Chapman, C., Allsop, S., Hides, L., . . . Birrell, L. (2014). The CLIMATE schools combined study: a cluster randomised controlled trial of a universal Internet-based prevention program for youth substance misuse, depression and anxiety. *BMC Psychiatry, 14*(1), 32.

Van De Looij, J., Petra, M., & De Wilde, E. J. (2008). Comparison of Web-Based versus Paper-and-Pencil Self-Administered Questionnaire: Effects on Health Indicators in Dutch Adolescents. *Health services research, 43*(5), 1708-1721.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*(4), 561-577.

Figure 1a

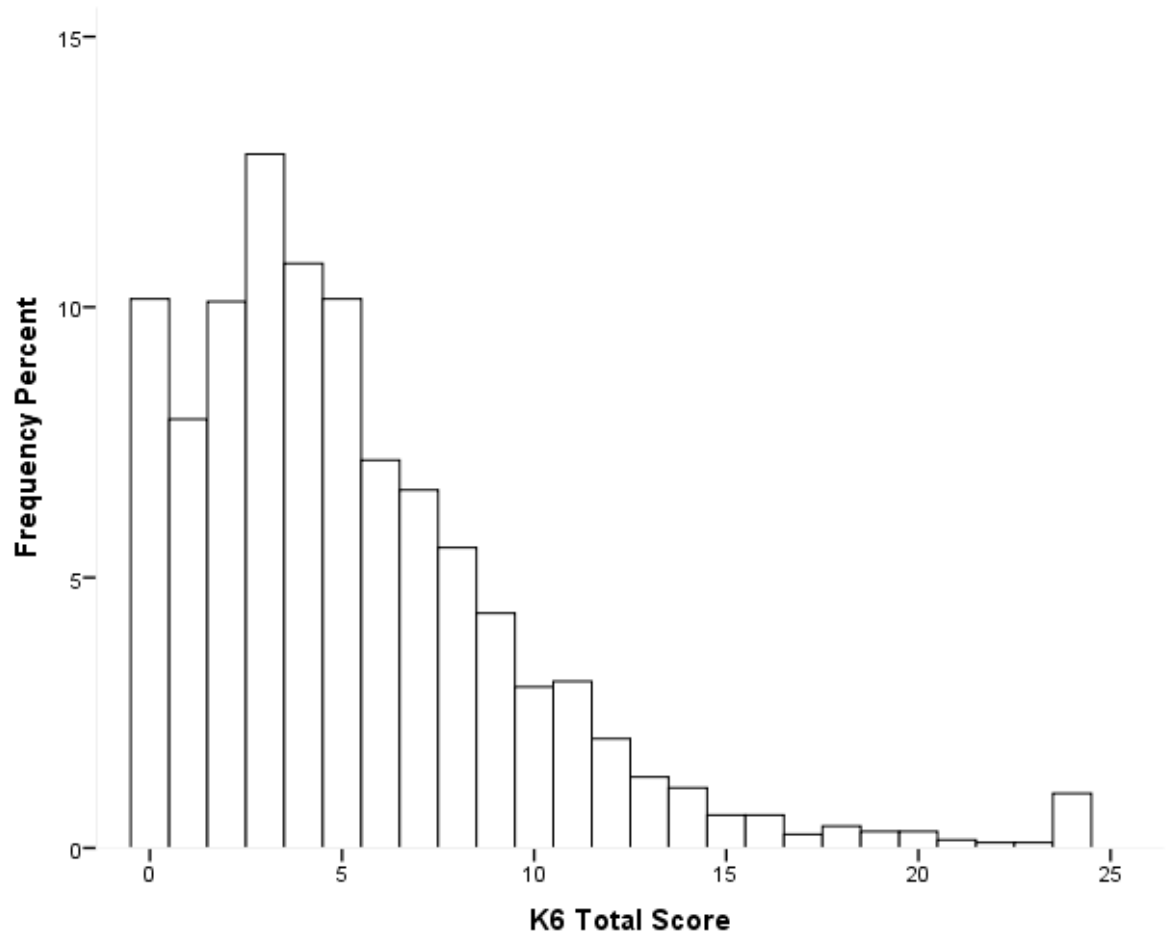The distribution of K6 scores amongst males in the [redacted for review] study ($N = 1,978$)

Frequency Percent vs K6 Total Score

Figure 1b

The distribution of K6 scores amongst females in the [redacted for review] study (*N* = 2,456)
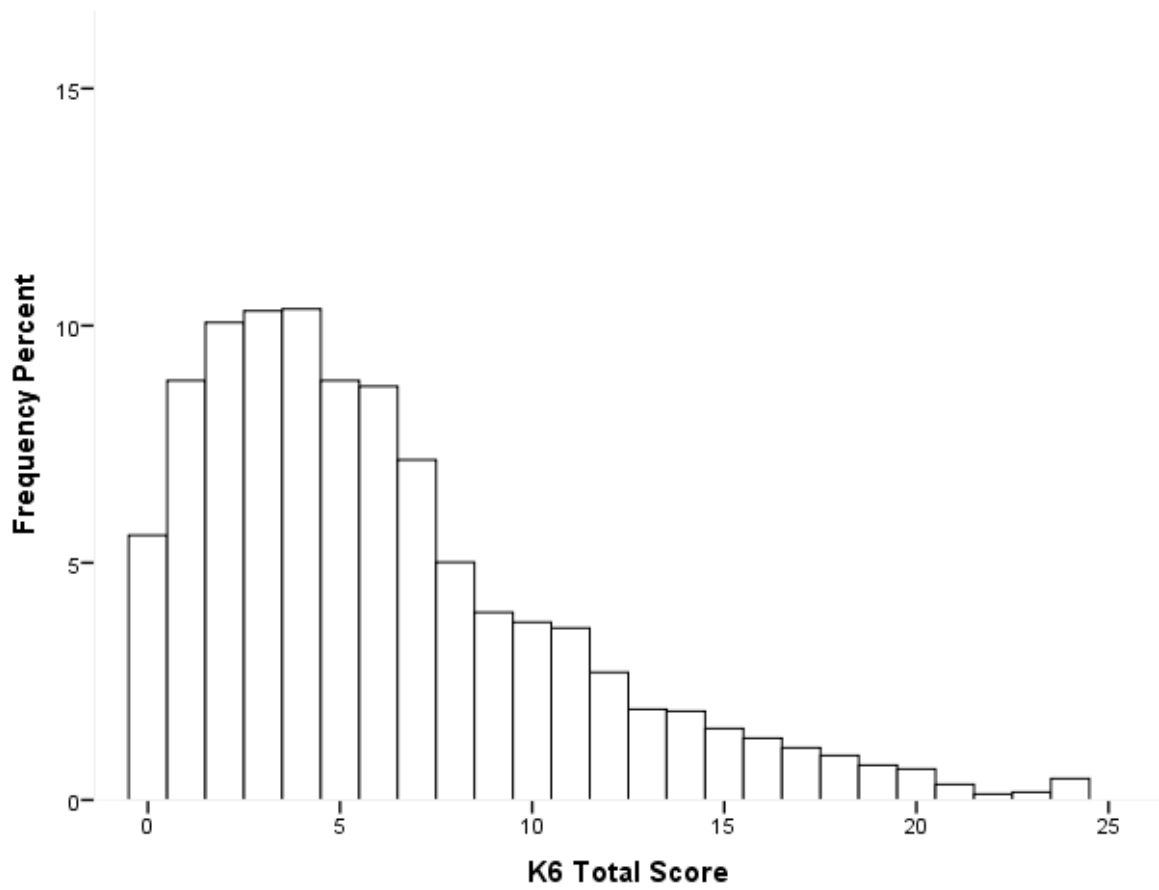
Figure 2

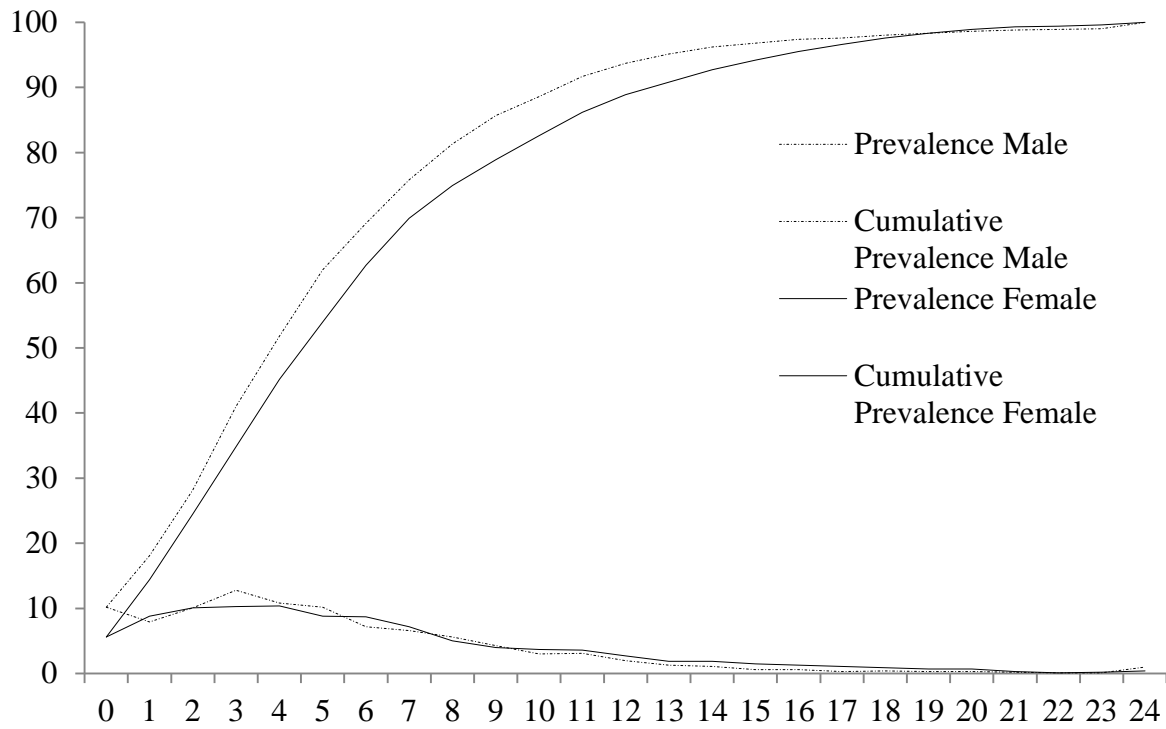Prevalence and cumulative prevalence of K6 scores in the [redacted for review] study (n = 4,434)

Figure 3

Distribution of each of the K6 items by sex in the [redacted for review] study (n = 4,434)
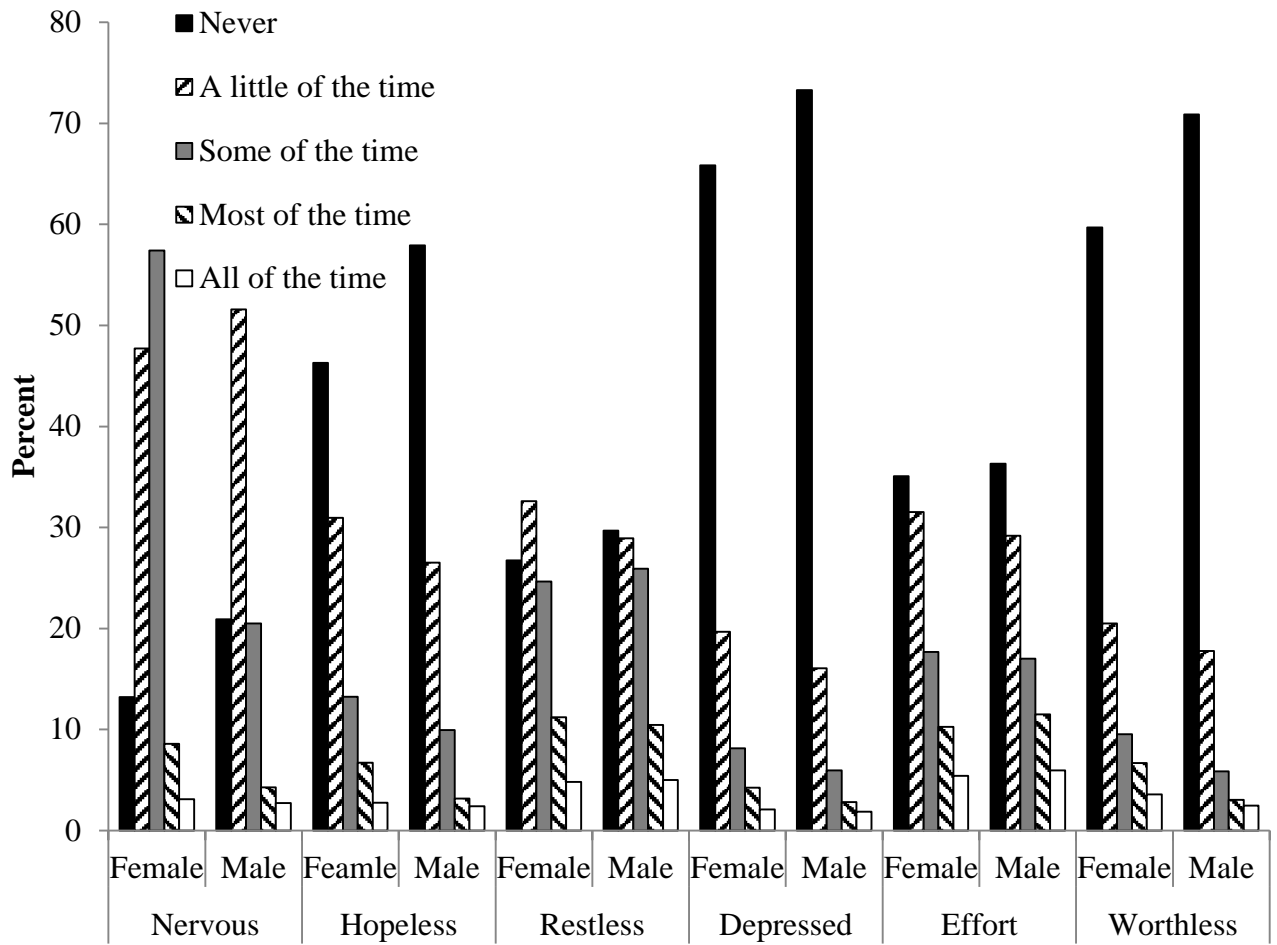
Figure 4a

Item response theory information functions for the K6 Nervous item in the [redacted for review] study (n = 4,434)
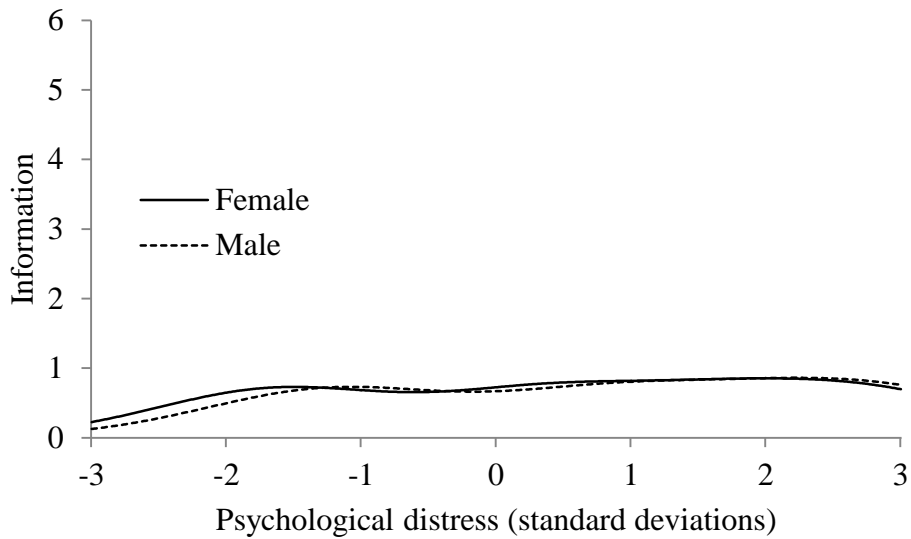

Figure 4b

Item response theory information functions for the K6 Hopeless item in the [redacted for review] study (n = 4,434)
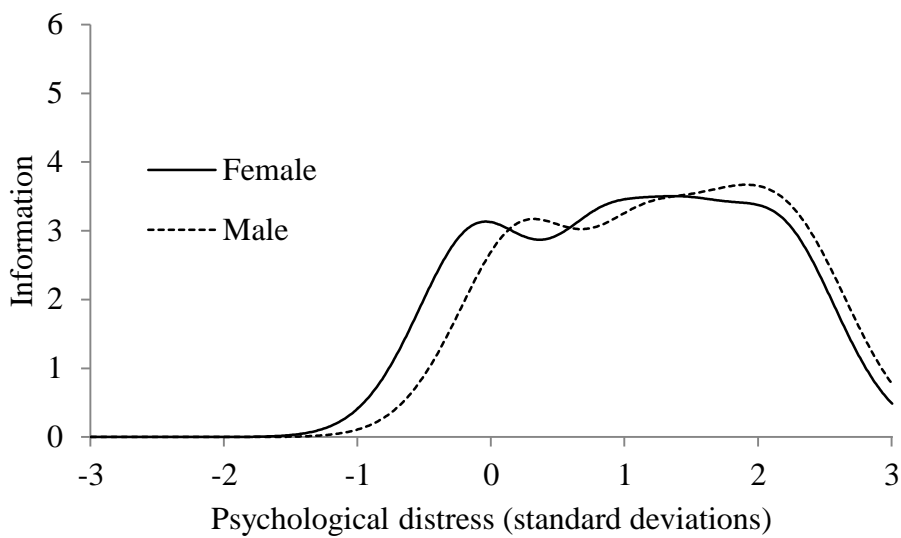
Figure 4c

Item response theory information functions for the K6 Restless item in the [redacted for review] study (n = 4,434)
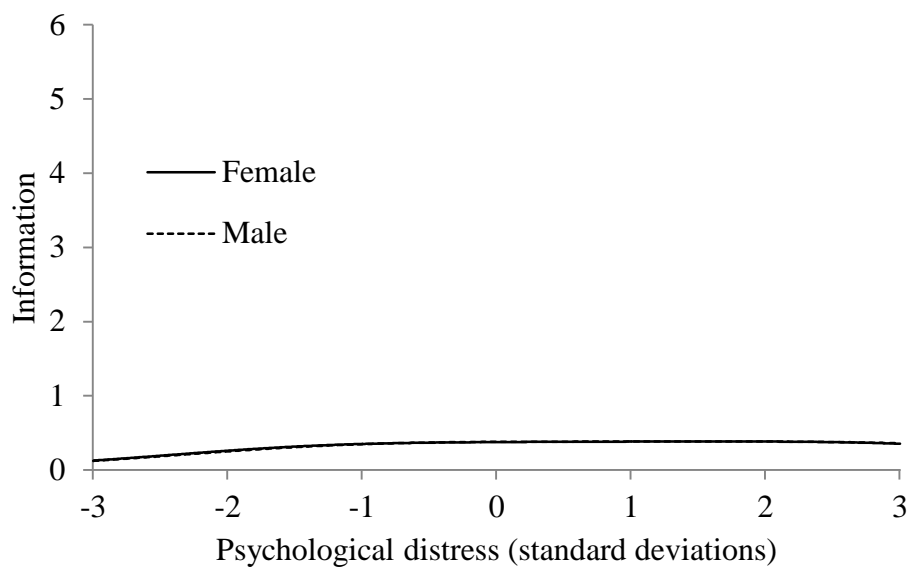

Figure 4d

Item response theory information functions for the K6 Depressed item in the [redacted for review] study (n = 4,434)
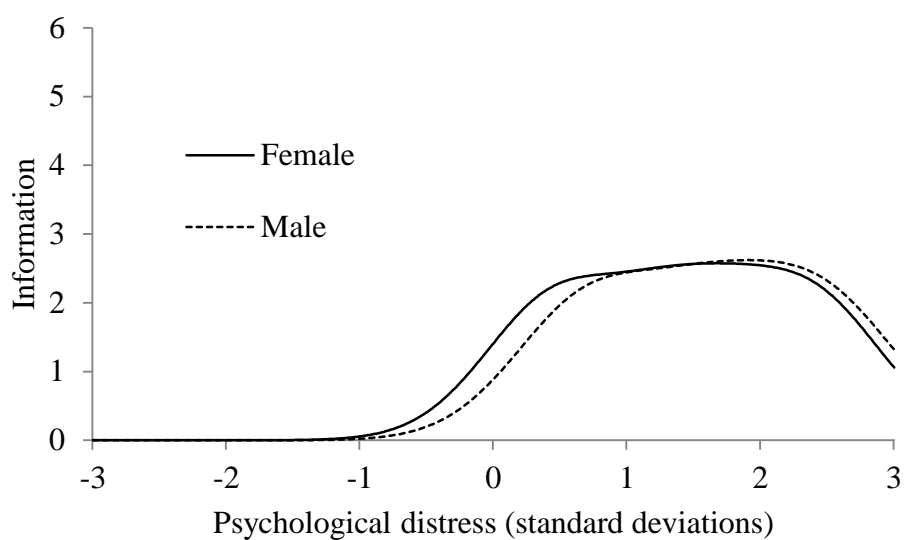
Figure 4e

Item response theory information functions for the K6 Effort item in the [redacted for review] study (n = 4,434)
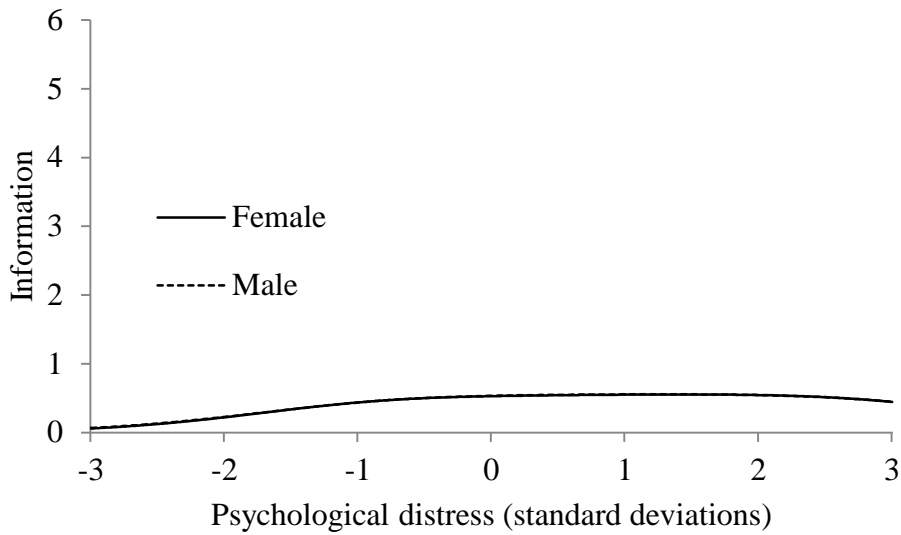

Figure 4f

Item response theory information functions for the K6 Worthless item in the [redacted for review] study (n = 4,434)
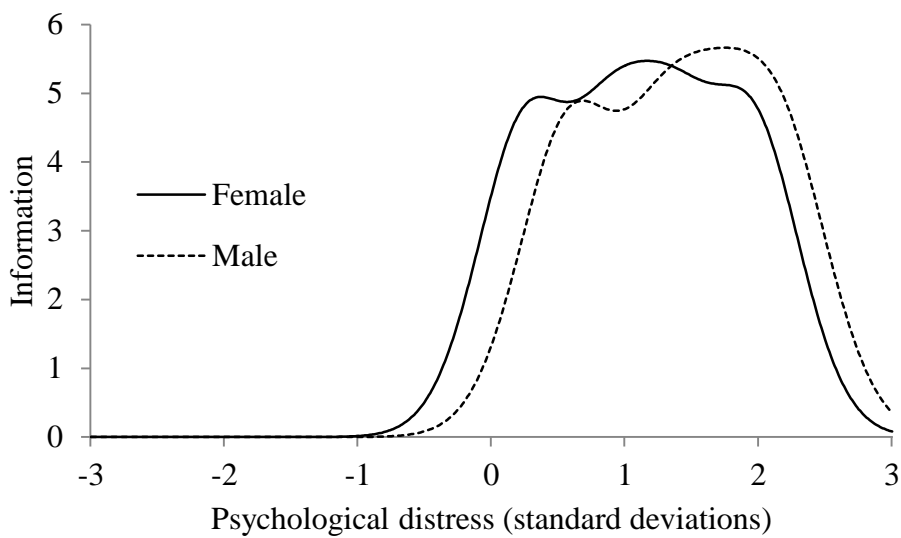
Figure 5

Item response theory test information function for the K6 scale in the [redacted for review]
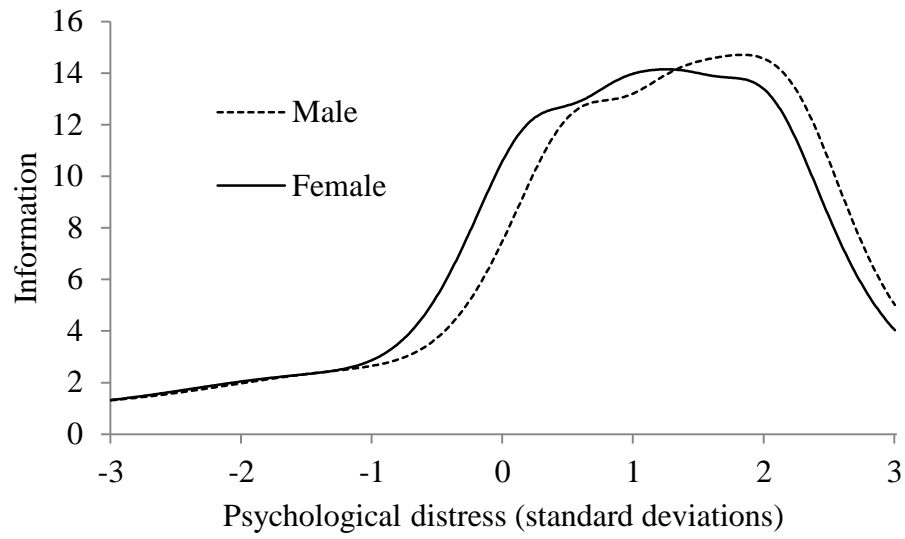
study (n = 4,434)

Table 1

K6 scores at selected percentiles of the K6 distribution by sex in the [redacted for review]

study (n = 4,434)

	Total sample (N = 4,434)	Males (N = 1,978)	Females (N = 2,456)
10th percentile	0	0	0
25th percentile	2	2	2
50th percentile	4	4	5
75th percentile	7	7	8
90th percentile	12	10	13
95th percentile	15	13	16
99th percentile	20	23	20
Mean (SD) K6 score	5.76 (4.67)	5.28 (4.44)	6.15 (4.82)

Table 2

Means, standard deviations, correlations, and factor loadings from exploratory factor analysis

of K6 items in the [redacted for review] study (n = 4,434) [1]

	Nervous	Hopeless	Restless	Depressed	Effort	Worthless
Nervous	1					
Hopeless	0.62	1				
Restless	0.49	0.50	1			
Depressed	0.55	0.76	0.43	1		
Effort	0.43	0.52	0.46	0.53	1	
Worthless	0.54	0.83	0.48	0.80	0.52	1
Mean	1.30	0.78	1.34	0.51	1.20	0.63
SD	0.92	1.01	1.14	0.92	1.20	1.03
Factor loadings (1-factor model)	0.67	0.90	0.60	0.85	0.62	0.91
Factor loadings (Factor 1, 2-factor model)	0.38	0.83	0.00	0.85	0.35	0.97
Factor loadings (Factor 2, 2-factor model)	0.38	0.10	0.78	0.00	0.34	-0.07

[1] The six eigenvalues from the exploratory factor analysis: 3.86, 0.70, 0.58, 0.46, 0.24, 0.16

Table 3

Fit statistics for the exploratory factor analysis (EFA) and confirmatory factor analyses

(CFA) in the [redacted for review] study (n = 4,434)

RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI:

	χ^2 $_{(df)}$	RMSEA	CFI	TLI
EFA				
1 factor EFA	530.532 $_{(9)}$	0.114	0.982	0.970
2 factor EFA	101.255 $_{(4)}$	0.074	0.997	0.988
1 Factor CFA				
Model 1	530.991 $_{(9)}$	0.114	0.982	0.970
Model 1a	51.170 $_{(5)}$	0.046	0.998	0.995
2 Factor CFA				
Model 2	212.912 $_{(8)}$	0.076	0.993	0.987
Measurement invariance testing				
Baseline	60.807 $_{(10)}$	0.048	0.998	0.995
Weak invariance	44.974 $_{(19)}$	0.025	0.999	0.999
Strong invariance	374.812 $_{(43)}$	0.059	0.988	0.992

Tucker Lewis Index; df: degrees of freedom

Table 4

Factor loadings and thresholds for each of the K6 items in the [redacted for review] study (n = 4,434)

	Factor loading		Threshold 1		Threshold 2		Threshold 3		Threshold 4	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Nervous	0.681	0.697	-0.809	-1.117	0.597	0.279	1.474	1.188	1.923	1.867
Hopeless	0.887	0.895	0.200	-0.093	1.013	0.748	1.589	1.311	1.973	1.915
Restless	0.523	0.541	-0.533	-0.621	0.218	0.236	1.017	0.994	1.645	1.664
Depressed	0.848	0.858	0.621	0.408	1.245	1.060	1.675	1.529	2.082	2.037
Effort	0.590	0.608	-0.350	-0.383	0.399	0.430	0.936	1.007	1.558	1.606
Worthless	0.922	0.928	0.549	0.246	1.208	0.848	1.597	1.266	1.964	1.802

NB Unstandardized loadings were constrained to be equal, standardized loading varied across sex due to freely estimated factor variances. Loadings and thresholds as estimated in the weak invariance model.

Table 5

Means (standard deviations) and prevalence of abnormal scores in each of the subscales of the Strengths and Difficulties Questionnaire (SDQ) in the [redacted for review] study (n = 4,434)

	Whole sample (N = 4,434)		Males (N = 1,978)		Females (N = 2,456)	
	Mean (SD)	Prevalence of abnormal score (%)	Mean (SD)	Prevalence of abnormal score (%)	Mean (SD)	Prevalence of abnormal score (%)
Emotion	2.33 (2.36)	6.7	1.85 (2.18)	4.5	2.72 (2.42)	8.6
Conduct	1.82 (1.77)	9.1	1.95 (1.80)	10.4	1.71 (1.73)	8.0
Hyperactivity	3.21 (2.22)	8.0	3.26 (2.17)	7.7	3.17 (2.27)	8.3
Peer Difficulties	1.96 (1.77)	4.2	2.13 (1.83)	4.9	1.82 (1.70)	3.6
Total Difficulties	9.31 (6.14)	7.9	9.18 (6.12)	8.4	9.42 (6.16)	7.5

Table 6

Area Under the Curve (AUC) for the K6 scale and each of the subscales of the Strengths and Difficulties Questionnaire in the [redacted for review] study (n = 4,434)

	Whole sample (N = 4,434)		Males (N = 1,978)		Females (N = 2,456)	
	AUC	95% CI	AUC	95% CI	AUC	95% CI
Emotion	0.856	0.844-0.867	0.783	0.764-0.802	0.891	0.876-0.905
Conduct	0.713	0.699-0.728	0.666	0.644-0.687	0.788	0.769-0.806
Hyperactivity	0.770	0.756-0.783	0.753	0.733-0.773	0.786	0.767-0.804
Peer Difficulties	0.718	0.704-0.733	0.682	0.661-0.703	0.783	0.764-0.802
Total Difficulties	0.793	0.780-0.806	0.719	0.698-0.739	0.883	0.867-0.897