This is the published version of:


Access to the published version:

http://doi.org/10.1002/mpr.1363

Copyright:

Copyright the Publisher 2012. Version archived for private and non-commercial use with the permission of the author/s and according to publisher conditions. For further rights please contact the publisher.
Are Male and Female Responses to Social Phobia Diagnostic Criteria Comparable?

Erica Crome a

Assoc. Prof. Andrew Baillie a

Dr Alan Taylor b

a. Centre for Emotional Health
   Psychology Department
   Macquarie University, NSW
   Australia

b. Psychology Department
   Macquarie University, NSW
   Australia

Corresponding Author:
Erica Crome
Psychology Department
Macquarie University NSW Australia 2109
PH: +61-2-9850-7941, FAX: +61-2-9850-8062,
Email: erica.crome@mq.edu.au
Abstract

Females typically report higher social phobia levels than males in community samples, and this may be due to sex bias in assessment measures. This study aims to establish whether patterns of responding to social phobia diagnostic criteria in the Composite International Diagnostic Interview (CIDI) are comparable across males and females. A subsample of participants in the Australian National Survey of Mental Health and Wellbeing (1997) reporting at least one social fear were selected (n = 1755). Analyses were conducted using a series of multi-group confirmatory factor analyses for categorical data, with unique steps to model invariance of residual variances. Partial, but not full, invariance was established, as males and females differed in their responses to items assessing physical anxiety symptoms at low levels of social fear. Whilst these differences were statistically significant, they are likely not to affect clinical practice or rates of social phobia diagnosis. This supports differences on this measure being interpreted as genuine, and strengthens findings females are more vulnerable to social phobia than males.

Key words: social phobia, measurement invariance, confirmatory factor analysis
1. Introduction

Social phobia, or social anxiety disorder, is characterised by anxiety regarding social situations and potential negative evaluation (American Psychiatric Association (APA), 2000). It is one of the most common mental disorders in epidemiological surveys, with estimates of lifetime prevalence in Western communities as high as 16 percent (Hidalgo et al., 2001). Social phobia typically emerges in adolescence and is associated with chronic distress, impairment and comorbid disorder (Furmark, 2002). Females in community samples are more likely than males to meet social phobia criteria, with differential neurobiology, socialisation and life experiences all presented as possible contributors to this difference (Rapee et al., 2004). However, before these avenues are researched further, a more basic explanation must be ruled out. That is, that differential sex prevalences are due to differences in how males and females respond to diagnostic measures. If such a measurement bias exists, it is possible males and females would appear differentially affected, despite actually having comparable levels of underlying social phobia. Therefore, it is important that, before male and female responses on any diagnostic measure can be meaningfully compared, this underlying measurement assumption must be tested.

There is growing evidence males and females do differ in their patterns of responses to psychological assessments. For example, females have a generally tendency to use extreme response options (e.g. “strongly agree”, “strongly disagree”) more frequently than males (Ritvo et al., 2008). There are also important sex differences in recall, willingness to self-report distress, interpretation of symptomatology, interviewer effects and treatment seeking behaviour (Grant et al., 2007). These measurement biases can even occur in diagnostic criteriia themselves, with
Agrawal et al. (2007) reporting the wording of “hazardous use” criteria for cannabis abuse biased diagnosis towards males. Whilst differences associated with measurement variance may appear trivial, potential consequences include misdiagnosis and inappropriate allocation of resources, as well as misdirecting research into etiology and treatment differences which may not exist (Wisner et al., 2007). Assessing the accuracy of assessment tools is especially timely given both international mental health assessment systems *International Statistical Classification of Diseases* (10th ed.: ICD-10: World Health Organisation (WHO), 1992) and *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. text revision: DSM-IV: APA, 2000) are under review (Bögels, 2010).

The assumption people from different groups respond to assessment items in the same way is referred to as measurement invariance. In order for groups, such as males and females, to be meaningfully compared this assumption must be met. If measures are not invariant, systematic differences between group responses may affect the comparability of outcomes, and group differences may be more indicative of another confounding factor (Meredith et al., 2006). An example of this can be found in Chung et al.’s (2008) finding that whether or not a person experienced an assultive or non-assultive trauma, in addition to their level of post-traumatic stress, determined self-reported distress and emotional numbing. Despite commonly researched sex differences in social phobia, there is minimal research supporting the invariance of common diagnostic measures. The most common measure of mental disorder prevalence is the Composite International Diagnostic Interview (CIDI: Kessler, Ustun et al., 2004; WHO, 1997). In Australia, variants of the CIDI have been used to assess the prevalence of mental disorders, assess the burden of disease and guide both mental health care policy and research (Andrews et al., 2004;
Issakidis et al., 2004). Given the gravity of outcomes from these surveys, it is essential any potential measurement biases are detected. This paper aims to test the underlying measurement invariance assumption of the social phobia diagnostic section of the CIDI used in Australian surveys to determine whether outcomes for males and females can be meaningfully compared.
2. Material and Methods

2.1 Participants:

Australian National Survey of Mental Health and Wellbeing (NSMHWB) 1997

The NSMHWB 1997 provides the most recent Australian data suitable for measurement invariance analyses of a social phobia diagnostic measure. It was conducted by the Australian Bureau of Statistics (ABS) within a multi-stage stratified random sample of adults in private dwellings representative of the Australian population. In total, 10,641 people aged 18 and over participated in the voluntary face-to-face survey (response rate of 78.1%: Andrews et al., 2001). Of 10,641 respondents, 1755 (16.5%) endorsed items related to an unusually strong fear and/or avoidance of at least one of seven possible social scenarios (e.g. giving a speech or speaking in public) and thus screened into the social phobia diagnostic section. These responders were used in current analyses. Corresponding population weights for this subset were used to increase generalisation to Australians experiencing social fears as per ABS Guidelines (1997). Sample demographics and proportions of symptom endorsement are detailed in Table 1.

INSERT TABLE 1 ABOUT HERE

2.2 Measure:

Composite International Diagnostic Interview (CIDI)

The NSMHWB 1997 survey was based on the CIDI v2.1 (WHO, 1997), a fully structured, standardised diagnostic interview covering both ICD-10 and DSM-IV diagnoses. The CIDI has
strong psychometric properties (Andrews et al., 1998; Pennell et al., 2004) and was administered by trained lay interviewers via computer-assisted personal interviews. The CIDI v2.1 assesses the 12 month prevalence of DSM-IV social phobia using diagnostic criteria used to structure analyses. These were: A. marked fear of social situations with potential for humiliation or embarrassment (fear), B. exposure to the situation invokes sensations of anxiety (physical anxiety), C. recognition the fear is unreasonable or excessive (excessive), D. feared situations are avoided or endured with distress (avoid / endure), E. social fears cause significant interference (interference) and G&H. social fears are not due to other conditions such as substance use, medical illness or another mental disorder (not due to other: APA, 2000). Criterion F was excluded as it applies to people under the age of 18 who were not included in the survey. Criteria were dichotomously coded as either “not present” or “present” and further information on questions and criteria determination can be found in Robins et al. (1989).

2.3 Statistical Analyses:
The measurement invariance hypothesis was explored using a series of multi-group confirmatory factor analyses (MGCFA). These analyses test whether differences in the pattern of male and female responses emerge when restrictions to model comparability are imposed at each step (Byrne, 2004; Steenkamp et al., 1998). This method was based on a two-step procedure recommended for categorical data outlined by Muthén et al. (2008). These models are based on the assumption that if a construct is invariant, all people experiencing similar levels of an underlying construct should report a similar pattern of symptoms. When using categorical data, the underlying construct is modelled by a continuous and normally distributed latent response variable, based on the assumption each categorical indicator could have been measured in a more
precise manner (Millsap et al., 2004). Firstly, a Baseline Model was created to model similarity in factor structure between males and females (freely estimated factor loadings and thresholds; Figure 1a). Then in step two, this model was compared to a second Measurement Invariance Model, modelling similarities between males and females on both the factor loading (how strongly each item related to underlying social phobia) and threshold (the level of social phobia at which a response changed from a “no” to a “yes”) of each item (Figure 1b). If there no significant differences between these two models, it is interpreted as a sign of no systematic differences in how males and females approach each assessment item. On the other hand, significant differences between the models suggests males and females systematically differ on some questionnaire items, whether this be due to real differences or measurement bias.

Whilst this two-step approach overcomes common modelling limitations associated with categorical data, it may ignore important sex differences in residual variance. Residual, or error, variance refers to the variance in each assessment item not accounted for by the underlying construct. For example, males and females may respond in an equivalent manner to items worded in a positive manner, yet differ in their reactions to negatively worded items, despite both assessing the same underlying construct (e.g. social phobia). Assessing potential differences in residual variance is important to ensure the accuracy of invariance analyses (Meredith et al., 2006; Steenkamp et al., 1998). To address this, a third step was included which compared an additional model, the Strict Invariance Model (Figure 1c), with the Measurement Invariance Model to explore equality of residual variance. If significant differences emerged between consecutive models at any step, the measurement invariance hypothesis would be rejected and partial invariance would be explored. This determines whether any items can be used to provide
meaningful comparisons between the two groups (Steenkamp et al., 1998). Modification indicies estimating changes in model fit with specific changes can be used to guide the formation of partial invariance models (Schreiber, 2008).

The current analyses were conducted in MPlus (v 5.21: Muthén et al., 2009) using a robust weighted least squares (WLSMV) estimator for categorical data. Theta parametization was selected to allow the specification of residual variances (Millsap et al., 2004; Muthén et al., 2008). Model fit was assessed via multiple fit indexes, including a robust chi-squared test ($\chi^2$) for WLSMV estimators, comparative fit index (CFI), Tucker-Lewis index (TFI) and root mean square error of approximation (RMSEA) and the weighted root mean square residual (WRMR: Brown, 2006; Schreiber, 2008). Non-significant $\chi^2$ values suggest model retention, and cut-off values for other fit indexes in categorical data were $\geq 0.95$ for CFI, $\geq 0.96$ for TFI, $< 0.90$ for WRMR and $< 0.06$ for RMSEA (Schreiber, 2008).

Two indexes were used to assess differences between the models, the first being difference in $\chi^2$ adjusted for nested models using MPlus DIFFTEST. Non-significant differences on this index supports the assumption of comparability between male and female responses, as adding increased restrictions did not significantly alter model fit (Muthén et al., 2008). Changes in CFI values were the second change index, with differences less than .01 suggesting no significant difference between models (Cheung et al., 2002). Current diagnostic conceptualizations of social phobia guided model specification, with all criteria loading onto a single social phobia factor. This was supported by evidence of an underlying social phobia dimension (Crome et al., 2010). Using MPlus defaults, the factor metric was set by the first indicator (A. fear: Muthén et al.,
and the baseline model was tested in males and females separately prior to analysis (Byrne, 2004). To conserve space, parameters from only the initial and final standardized solutions are reported and other estimates are available on request.

INSERT FIGURES 1A, B AND C ABOUT HERE
3. Results

The one factor model appeared more applicable to females; however, fit indexes (detailed above) largely supported combining male and female responses to create the Baseline Model (Table 2). Fit indexes for the Baseline Model supported model retention, so this model was then compared to the Measurement Invariance Model in step two.

Change of fit indexes provided mixed support for equivalence of these two models, with $\Delta \chi^2$, but not $\Delta$ CFI, reflecting significant differences between models. Evidence for the increased ability of $\chi^2$ over CFI to detect significant differences between models guided decisions to reject the more restrictive model and thus the measurement invariance hypothesis (French & Finch, 2006). As full measurement invariance was not established, partial invariance was explored. Modification indices suggested model improvement if males and females were allowed to differ in their report of physical anxiety symptoms (B. physical anxiety) and physical anxiety symptoms were allowed to vary as a function of social fears (A. fear) for males only. These relationships were adopted to form the revised Partial Invariance Model detailed in Figure 2.

Whilst not directly compared, these modifications appeared to provide a better picture of the performance of diagnostic items for the overall sample. As seen by $\Delta \chi^2$ and $\Delta$ CFI in Table 2, increasing model restrictions in step two (Partial Invariance Model) or step three (Strict
Measurement Invariance Model) did not significantly affect model fit. This provides support for invariance between male and female responses on all diagnostic criteria aside from B. exposure to the situation invokes sensations of anxiety (physical anxiety). This difference was explored using logistic regression and cross-tabulation in SPSS (2006). Logistic regression identified a significant interaction between sex and social fears in the report of physical anxiety symptoms (interaction $\beta = -0.575 (0.210), p < .05$: Males = 0, Females = 1).

Cross-tabulation clarified this interaction, with evidence of minimal difference in the proportion of males and females reporting physical anxiety symptoms in the presence of a significantly high social fear (70.4% & 73.9% respectively), yet a much higher proportions of females (73.9%) over males (32.8%) reporting physical anxiety symptoms at low levels of social fear. This highlighted differences in the specificity (number of people without social phobia who also report no physical symptoms) of this item for males and females. A higher specificity for males (0.55: 95% C.I. = 0.51 to 0.58) compared with females (0.43: 95% C.I. = 0.40 to 0.47: $\chi^2 = 10.233 (1), p = .001$) meaning this item may be more informative of actual social phobia status for males. Differences in the sensitivity (true positives) could not be explored as every diagnosis of social phobia requires this criterion to be met. The clinical significance of this variance is unclear as, in practice, social phobia criteria are rarely considered in isolation. This effect on diagnosis would also only be observed in the 1.5% of females and 1.4% of males in the sample meeting all other diagnostic criteria aside from criterion B.
4. Discussion

The social phobia diagnostic section of the CIDI used in the NSMHWB 1997 meets criteria for partial measurement invariance. This means people with the same level of underlying social phobia will respond to most diagnostic items in comparable manner, regardless of whether they are male or female. Full invariance was not established as males tended to report fewer physical anxiety symptoms than females at low social fear levels. Whilst Criterion B (physical anxiety) appears more informative for males, it is unlikely this has any real significance for clinicians and researchers as it is unlikely to affect rates of diagnosis. Even within the small populations of males and females with potential for misdiagnosis based on this one criterion, given they also meet all other diagnostic criteria, they are unlikely to also experience the low levels of social fears where this relationship is observed. Overall, findings of strong partial invariance suggests the higher proportion of females in this sample being classified as having social phobia (Lampe et al., 2003) reflects a genuine difference in the experience and incidence of social phobia between males and females.

This finding is consistent with an emerging measurement invariance literature demonstrating comparability between male and female responses to other clinical social anxiety measures including the Social Phobia and Anxiety Inventory (SPAI: Roberson-Nay et al., 2007) and Goldberg Anxiety and Depression Scales (Leach et al., 2008). The precise cause of the sex difference in physical symptoms at low levels of social anxiety is unclear, however it is consistent with findings socially anxious females are hyper-responsive on self-report, hemodynamic and autonomic measures. In contrast, males tend to display minimal physiological changes despite self-reported psychological symptoms (Grossman et al., 2001). There is also
evidence that the experience of social phobia may be qualitatively different for males and females based on factors not assessed during diagnosis, such as age of onset, severity of initial symptoms and rates of comorbidity (Bezerra De Menezes et al., 2008). Subtle differences in the report of physical anxiety symptoms may be one of these differences actually captured during diagnosis. Alternatively, differences in physical symptoms may be due to findings females generally report more general physical symptoms regardless of pathology, measure or response format used or population sampled (Barsky et al., 2001; Gijsbers van Wijk et al., 1999).

Interestingly, there is also a large overlap between general physical symptoms reported by females in community samples and those used to assess anxiety or distress (e.g. fatigue, dizziness, difficulty concentrating and sleeping, nausea and dyspnoea: Barsky et al., 2001). Differences in socialisation may also contribute to this difference, with males typically encouraged to rely on other forms of expression such as anger, over involvement in work or isolation rather than reporting symptoms of anxiety or distress (Rochlen et al., 2010).

This research has many strengths, including using a large epidemiological sample, and methods and software best suited to the available data. Few researchers have applied Muthen’s methods (e.g. Batinic et al., 2008; Hutchinson et al., 2008; Leach et al., 2008) despite the potential for erroneous results associated with applying common measurement invariance methods designed for continuous data to ordinal or Likert-type data (Schreiber, 2008; Brown, 2006). This research has the additional strength of extending Muthen’s methods to test invariance in residual variances, claimed to be essential for fair comparisons between groups (Meredith et al., 2006). Whilst using retrospective self-report data is typically a limitation to psychiatric research, as patterns of responding were of interest in this study, it is not a concern in the current paper.
Similarly, whilst no data is available for participants reporting no social fears given they screened out of the diagnostic section, as this measure is not used for assessment in this population, relying on a subsample will not likely affect generalization.

Using categorical data within CFA does have limitations, such as reduced accuracy of fit indexes (French et al., 2006) and difficulty ensuring relationships between the latent factor and latent response variable extend to observed variables (Millsap et al., 2004). However, the impact of this is likely to be reduced given the current sample was large and the model tested was simple (French et al., 2006). These analyses also do not provide a definitive answer as to whether any group differences are due to measurement bias or real differences in the experience of that particular symptom. Other concerns about error in hierarchical rules for diagnosis in the CIDI v2.1 social phobia section (Means-Christensen et al., 2003) did not apply to the current study, as diagnostic criteria were assessed without applying hierarchy rules. This also means sex differences may still emerge if exclusion criteria differentially affect the sexes. The CIDI used in the NSMHWB 1997 has a narrower range of social situations used as screening items than versions of the CIDI used in more recent WHO World Mental Health (WMH) surveys (Kessler, Ustun et al., 2004). This potentially means not all participants with social fears screened into the diagnostic section.

Whilst this study assesses measurement invariance in the diagnostic criteria of the CIDI, it is important to acknowledge this does not necessarily precisely extend to social phobia diagnostic criteria prescribed by the DSM-IV. For example, relying primarily on physical anxiety symptoms may not fully represent criterion B (*exposure to the feared social situation almost invariably*
provokes anxiety, which may take the form of a situationally bound or situationally predisposed panic attack; pg. 456, APA, 2000). It may be physical symptoms are non-specific markers of distress rather than specific symptoms of social phobia (Baillie et al., 2005; Clark et al., 1994). The opportunity to explore differences in the report of physical symptoms between males and females is further limited by screening points within the CIDI designed to reduce participant load. Once participants report a sufficient number of physical anxiety symptoms to meet criterion B, no further physical anxiety symptoms are elicited, which creates a potential ceiling effect in the data.

Overall this research supports the invariance of the social phobia diagnostic section of the CIDI used in the NSMHWB 1997. This strengthens previous findings Australian females are more vulnerable to social phobia and supports research into differential causal and maintaining factors (Grant et al., 2007). This research also identifies differences in how males and females may respond to items regarding physical anxiety symptoms at low levels of social fears. However, this difference does not appear likely to affect rates of social phobia diagnosis or responses of people experiencing clinical levels of social phobia. This research establishes the importance of explicitly testing underlying measurement invariance assumptions prior to interpreting group differences. Comparable analyses should be conducted in other versions of the CIDI, such as that used in the National Comorbidity Survey- Replication (Kessler, Berglund et al., 2004), to confirm that slight variations between surveys do not affect measurement invariance.
Table 1: Demographics and diagnostic status of the National Survey of Mental Health and Wellbeing (1997) participants screening into the social phobia section of the Composite International Diagnostic Interview (n = 1755).

<table>
<thead>
<tr>
<th></th>
<th>Males (n = 733)</th>
<th>%</th>
<th>Females (n = 1022)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 – 35</td>
<td>283</td>
<td>38.6</td>
<td>388</td>
<td>38.0</td>
</tr>
<tr>
<td>35 – 55</td>
<td>307</td>
<td>41.9</td>
<td>476</td>
<td>46.6</td>
</tr>
<tr>
<td>55 to &gt;75</td>
<td>143</td>
<td>19.5</td>
<td>158</td>
<td>15.4</td>
</tr>
</tbody>
</table>

**Marital Status**

<table>
<thead>
<tr>
<th></th>
<th>Males (n = 733)</th>
<th>%</th>
<th>Females (n = 1022)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married / Defacto</td>
<td>383</td>
<td>52.3</td>
<td>553</td>
<td>54.1</td>
</tr>
<tr>
<td>Separated / Divorced / Widowed</td>
<td>111</td>
<td>15.1</td>
<td>261</td>
<td>25.5</td>
</tr>
<tr>
<td>Never Married</td>
<td>239</td>
<td>32.6</td>
<td>208</td>
<td>20.4</td>
</tr>
</tbody>
</table>

**Employment**

<table>
<thead>
<tr>
<th></th>
<th>Males (n = 733)</th>
<th>%</th>
<th>Females (n = 1022)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-time</td>
<td>472</td>
<td>64.4</td>
<td>306</td>
<td>29.9</td>
</tr>
<tr>
<td>Part-time</td>
<td>64</td>
<td>8.7</td>
<td>298</td>
<td>29.2</td>
</tr>
<tr>
<td>Unemployed</td>
<td>45</td>
<td>6.1</td>
<td>60</td>
<td>5.9</td>
</tr>
<tr>
<td>Not in labour force</td>
<td>152</td>
<td>20.8</td>
<td>358</td>
<td>35.0</td>
</tr>
</tbody>
</table>

**Primary Language**

<table>
<thead>
<tr>
<th></th>
<th>Males (n = 733)</th>
<th>%</th>
<th>Females (n = 1022)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>706</td>
<td>96.3</td>
<td>980</td>
<td>95.9</td>
</tr>
<tr>
<td>Other</td>
<td>27</td>
<td>3.7</td>
<td>42</td>
<td>4.1</td>
</tr>
</tbody>
</table>

**DSM-IV Diagnosis**

<table>
<thead>
<tr>
<th></th>
<th>Males (n = 733)</th>
<th>%</th>
<th>Females (n = 1022)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>12mth Social Phobia</td>
<td>62</td>
<td>8.5</td>
<td>83</td>
<td>8.1</td>
</tr>
<tr>
<td>A. Fear present</td>
<td>352</td>
<td>44.3</td>
<td>443</td>
<td>55.7</td>
</tr>
<tr>
<td>B. Physical anxiety present</td>
<td>401</td>
<td>41.6</td>
<td>563</td>
<td>58.4</td>
</tr>
<tr>
<td>C. Excessive present</td>
<td>413</td>
<td>44.9</td>
<td>507</td>
<td>55.1</td>
</tr>
<tr>
<td>D. Avoid / endure present</td>
<td>422</td>
<td>44.7</td>
<td>523</td>
<td>55.3</td>
</tr>
<tr>
<td>E. Interference present</td>
<td>351</td>
<td>42.6</td>
<td>472</td>
<td>57.4</td>
</tr>
<tr>
<td>G&amp;H. Not due to other present</td>
<td>163</td>
<td>42.3</td>
<td>222</td>
<td>57.7</td>
</tr>
</tbody>
</table>

DSM-IV Criteria are: A. marked fear of social situations with potential for humiliation or embarrassment (*fear*); B. exposure to the situation invokes sensations of anxiety (*physical anxiety*); C. recognition the fear is unreasonable or excessive (*excessive*); D. feared situations are avoided or endured with distress (*avoid / endure*); E. social fears cause significant interference (*interference*); G&H. social fears are not due to other conditions such as substance use, medical illness or another mental disorder (*not due to other*).
Table Two: Fit indexes and change in fit indexes between models for individual groups, full measurement invariance and partial invariance analyses.

<table>
<thead>
<tr>
<th></th>
<th>$\chi^2$</th>
<th>df</th>
<th>$\Delta\chi^2$</th>
<th>df</th>
<th>RMSEA</th>
<th>WRMR</th>
<th>TLI</th>
<th>CFI</th>
<th>$\Delta$CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual Group Analyses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females (n = 1022)</td>
<td>10.743</td>
<td>9</td>
<td>.014</td>
<td>.576</td>
<td>.998</td>
<td>.998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males (n = 733)</td>
<td>23.417</td>
<td>8</td>
<td>.051</td>
<td>.938</td>
<td>.968</td>
<td>.974</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Measurement Invariance Analyses (Figure 1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline model</td>
<td>33.499</td>
<td>16</td>
<td>.035</td>
<td>1.101</td>
<td>.986</td>
<td>.989</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement Invariance Model</td>
<td>43.433</td>
<td>20</td>
<td>9.95</td>
<td>1.252</td>
<td>.985</td>
<td>.985</td>
<td>.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Partial Invariance Analyses (Figure 2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original Baseline Model</td>
<td>33.499</td>
<td>16</td>
<td>.035</td>
<td>1.101</td>
<td>.986</td>
<td>.989</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial Invariance Baseline Model</td>
<td>17.827</td>
<td>15</td>
<td>.015</td>
<td>.800</td>
<td>.998</td>
<td>.998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial Measurement Invariance Model</td>
<td>19.065</td>
<td>18</td>
<td>.008</td>
<td>.825</td>
<td>.999</td>
<td>.999</td>
<td>-.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stricter Partial Measurement Invariance Model</td>
<td>18.224</td>
<td>22</td>
<td>1.062</td>
<td>.856</td>
<td>1.002</td>
<td>1</td>
<td>-.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

($\chi^2$) chi-square goodness of fit index; (df) degrees of freedom; ($\Delta\chi^2$) adjusted change in $\chi^2$ for nested models; (RMSEA) root mean square error of approximation; (WRMR) weighted root mean square residual; (TFI) Tucker-Lewis Index; (CFI) comparative fit index; ($\Delta$CFI) change in CFI.

Figures in bold reflect significance at p < .05; non-significance (p = > .05) indicates better model fit.
Table Three: Standardised factor loadings, thresholds and variance explained ($R^2$) for Baseline and Strict Measurement Invariance models presented separately for males and females.

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th></th>
<th>Females</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standardised Factor Loading (S.E.)</td>
<td>Standardised Threshold (S.E.)</td>
<td>$R^2$ (S.E.)</td>
<td>Standardised Factor Loading (S.E.)</td>
</tr>
<tr>
<td><strong>Initial Multi-Group Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Fear †</td>
<td>.788 (.095)</td>
<td>.232 (.068)</td>
<td>.383 (.057)</td>
<td>.617 (.071)</td>
</tr>
<tr>
<td>B. Physical Anxiety</td>
<td>.807 (.105)</td>
<td>.040 (.068)</td>
<td>.394 (.062)</td>
<td>.637 (.075)</td>
</tr>
<tr>
<td>C. Excessive</td>
<td>1.101 (.141)</td>
<td>-.006 (.079)</td>
<td>.548 (.063)</td>
<td>1.143 (.125)</td>
</tr>
<tr>
<td>D. Avoid / Endure</td>
<td>.708 (.087)</td>
<td>-.040 (.065)</td>
<td>.334 (.055)</td>
<td>.747 (.083)</td>
</tr>
<tr>
<td>E. Interference</td>
<td>1.094 (.142)</td>
<td>.276 (.082)</td>
<td>.545 (.064)</td>
<td>1.289 (.141)</td>
</tr>
<tr>
<td>G &amp; H Not Due to Other</td>
<td>1.112 (.173)</td>
<td>1.268 (.143)</td>
<td>.553 (.077)</td>
<td>1.343 (.164)</td>
</tr>
<tr>
<td><strong>Final Partial Measurement Invariance Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Fear †</td>
<td>.618 (.063)</td>
<td>.203 (.047)</td>
<td>.276 (.041)</td>
<td>.636 (.059)</td>
</tr>
<tr>
<td>B. Physical Anxiety</td>
<td>.658 (.091)</td>
<td>.037 (.063)</td>
<td>.302 (.058)</td>
<td>.639 (.076)</td>
</tr>
<tr>
<td>C. Excessive</td>
<td>1.127 (.107)</td>
<td>.029 (.067)</td>
<td>.559 (.047)</td>
<td>1.160 (.106)</td>
</tr>
<tr>
<td>D. Avoid / Endure</td>
<td>.727 (.070)</td>
<td>-.045 (.051)</td>
<td>.346 (.043)</td>
<td>.749 (.067)</td>
</tr>
<tr>
<td>E. Interference</td>
<td>1.222 (.123)</td>
<td>.248 (.074)</td>
<td>.599 (.048)</td>
<td>1.259 (.113)</td>
</tr>
<tr>
<td>G &amp; H Not Due to Other</td>
<td>1.251 (.142)</td>
<td>1.390 (.117)</td>
<td>.610 (.054)</td>
<td>1.288 (.128)</td>
</tr>
</tbody>
</table>

Figures in bold reflect significance at $p < .05$
† Marker variable
Figure Legends

**Figure 1** Three step measurement invariance analysis based on a one factor model of social phobia: a) Step One (Baseline Model) with factor loadings, thresholds and residual variance freely estimated between males and females; b) Step Two (Measurement Invariance Model) factor loadings and thresholds restrained to equality; c) Step Three (Strict Measurement Invariance Model) residual variance restrained to equality.

**Figure 2 Final** Partial Measurement Invariance Model based on a one factor model of social phobia. Factor loadings and thresholds for criterion B (*physical anxiety*) freely estimated for males and females; residual variance between criteria A (*fear*) and B (*physical anxiety*) allowed to co-vary for males only.
Acknowledgements

Both National Survey of Mental Health and Well Being were funded by the National Health Branch of the Commonwealth Department of Health and Aged Care, under the National Mental Health Strategy. They were conducted by the Australian Bureau of Statistics. The current study was supported by funding Macquarie University Research Excellence Scholarship to the first author. We also acknowledge the productive feedback provided by four anonymous reviewers.

Competing interests: the authors have no competing interests.
References


Figure 1a

Figure 1

A

- A. Fear
- B. Physical
- C. Excessive
- D. Avoid/Endure
- E. Interfere
- G/H. Not Other

→ Male
→ Female

Social Phobia

Figure 1b

B

- A. Fear
- B. Physical
- C. Excessive
- D. Avoid/Endure
- E. Interfere
- G/H. Not Other

→ Male
→ Female
→ All
Figure 1c

Figure 2

Figure 2