

PLEASE NOTE:

This is the author's version of the manuscript accepted for publication in *Human Reproduction*. Changes resulting from the publishing process, namely editing, corrections, final formatting for printed or online publication, and other modifications resulting from quality control procedures, may have been subsequently added.

The published version can be found in: Sobral, M. P., Costa, M. E., Schmidt, L., & Martins, M. V. (2017). COMPI Fertility Problem Stress Scales is a brief, valid and reliable tool for assessing stress in patients seeking treatment. *Human Reproduction*, 32(2), 375-382. doi: 10.1093/humrep/dew315

**COMPI Fertility Problem Stress Scales is a brief, valid and reliable tool for
assessing stress in patients seeking treatment¹**

Running title: COMPI Fertility Stress Scales validation

MP Sobral¹, ME Costa^{1,2}, Lone Schmidt³, MV Martins^{1,2, *}

¹Faculty of Psychology and Education Sciences, University of Porto, 4200-135 Porto, Portugal; ²Center for Psychology at University of Porto, 4200-135 Porto, Portugal;

³University of Copenhagen, 1014 Copenhagen K, Denmark;

***Correspondence address:** Mariana V. Martins, Faculty of Psychology and Education Sciences, Porto University, R. Alfredo Allen, 4200-135 Porto, Portugal. E-mail: mmartins@fpce.up.pt

Abstract

Study question Are the Copenhagen Multi-Centre Psychosocial Infertility research program Fertility Problem Stress Scales (COMPI-FPSS) a reliable and valid measure across gender and culture?

Summary answer The COMPI-FPSS is a valid and reliable measure, presenting excellent or good fit in the majority of the analyzed countries, and demonstrating full invariance across genders and partial invariance across cultures.

What is known already Cross-cultural and gender validation is needed to consider a measure as standard care within fertility. The present study is the first attempting to establish comparability of fertility-related stress across genders and countries.

Study design, size, duration Cross-sectional study. First, we tested the structure of the COMPI-FPSS. Then, reliability and validity (convergent and discriminant) were examined for the final model. Finally, measurement invariance both across genders and cultures was tested.

Participants/materials, setting, methods Our final sample had 3923 fertility patients (1691 men and 2232 women) recruited in clinical settings from seven different countries: Denmark, China, Croatia, Germany, Greece, Hungary, and Sweden. Participants had a mean age of 34 years and the majority (84%) were childless.

Main results and the role of chance Findings confirmed the original three-factor structure of the COMPI-FPSS, although suggesting a shortened measurement model using less items that fitted the data better than the full version model. While data from the Chinese and Croatian subsamples did not fit, all other countries presented good fit (χ^2/df

≤ 5.4 ; comparative fit index ≥ 0.94 ; root-mean-square error of approximation ≤ 0.07 ; modified expected cross-validation index ≤ 0.77). In general, reliability, convergent validity, and discriminant validity were observed in all subscales from each country (composite reliability ≥ 0.63 ; average variance extracted ≥ 0.38 ; squared correlation ≥ 0.13). Full invariance was established across genders, and partial invariance was demonstrated across countries.

Limitations, reasons for caution Generalizability regarding the validation of the COMPI-FPSS cannot be made regarding infertile individuals not seeking treatment, or non-European patients. This study did not investigate predictive validity, and hence the capability of this instrument in detecting changes in fertility-specific adjustment over time and predicting the psychological impact needs to be established in future research.

Wider implications of the findings Besides extending knowledge on the psychometric properties of one of the most used fertility stress questionnaire, this study demonstrates both research and clinical usefulness of the COMPI-FPSS.

Study funding/competing interest(s) This study was supported by European Union Funds (FEDER/COMPETE – Operational Competitiveness Program), by national funds (FCT – Portuguese Foundation for Science and Technology) under the projects PTDC/MHC-PSC/4195/2012 and SFRH/BPD/85789/2012. There are no conflicts of interest to declare.

Trial registration number N/A

Key-words: fertility stress, psychometric properties, cross-cultural comparison, reliability and validity, measurement invariance

Introduction

The assessment of the psychosocial consequences of infertility has gained growing ground among researchers over the last decades. Initially, quantitative assessment of the impact of infertility was conducted through general self-report measures of psychological adjustment, such as stress, anxiety, or depression (e.g., Bernstein, Potts, & Mattox, 1985; Connolly, Edelman, Cooke, & Robson, 1992; Emery et al., 2003). However, because these measures failed to capture the distress arising from experiencing infertility and its treatments (L. Schmidt, 2009), questionnaires specifically targeting infertile samples were developed in order to accurately assess the psychosocial outcomes of this experience.

One of the most known measures of fertility-related stress is the Copenhagen Multi-Centre Psychosocial Infertility (COMPI)-Fertility Problem Stress Scales (COMPI-FPPS) (L. Schmidt, 1996; L. Schmidt, B. E. Holstein, U. Christensen, & J. Boivin, 2005). The COMPI-FPSS was developed based on the Fertility Problem Stress Inventory (Abbey, Andrews, & Halman, 1991) and interviews with patients (L. Schmidt, 1996) assessing the impact of infertility on the personal, social and marital domains of an individual's life. Since then, these scales have been widely used (e.g., Kagami et al., 2012; Lykeridou et al., 2011; Mariana V Martins, Costa, Peterson, Costa, & Schmidt, 2014; Passet-Wittig et al., 2014; Pinborg, Hougaard, Andersen, Molbo, & Schmidt, 2009; L Schmidt et al., 2003; L. Schmidt, B.E. Holstein, U. Christensen, & J. Boivin, 2005), examining for example infertility distress risk and protective factors (Mariana V. Martins et al., 2013), or gender differences (B. D. Peterson, Pirritano, Christensen, & Schmidt, 2008).

This broad recognition and intensive use led to some effort toward validation beyond internal consistency. Schmidt and colleagues (2003) ran exploratory factor

analysis in a sample of 2500 patients seeking treatment and three factors were extracted, corresponding to fertility stress in personal, marital and social domains. Later, Martins et al. (2013) confirmed this original 3-factor structure in a Portuguese sample, with the hypothesized model showing good fit to the observed data.. However, extensive cross-national and cross-gender validation of questionnaires using large samples are required in order to use a measure as standard care within fertility (van Empel et al., 2010). Additionally, to compare countries or genders, comparability of constructs needs to be ascertained. Even though measures administered to both genders and/or derived from cross-cultural adaptations (i.e., translations and preliminary statistical analyses) are often assumed as equivalent (Beaton, Bombardier, Guillemin, & Ferraz, 2000; Epstein, Santo, & Guillemin, 2015), comparability implies constructs to be shown as adequate and invariant (fully or at least partially) across the groups under consideration (Fontaine & Fischer, 2010; Steinmetz, 2011). The assessment of measurement invariance determines if participants across groups perceive the questions and underlying factors in the same way (Vandenberg and Lance, 2000), being a pre-requisite for conclusions on group differences. In other words, measurement invariance has to be assessed to prove that assumptions on significant group differences such as the well-known greater impact of infertility in women when compared to their male counterparts (see B. Peterson et al., 2012, for a review on gender differences) can be made. Otherwise, we still do not know whether those differences are due to different ways of experiencing infertility or different ways of self-reporting and interpreting items.

Cross-national evidence is particularly important because globalization brings the need to establish whether psychological theories can be transferred into other cultural settings (Sanchez, Spector, & Cooper, 2006), and it allows health professionals to assess each patient in the corresponding cultural context (Uysal-Bozkir, Parlevliet, &

de Rooij, 2013). To reproductive health researchers and clinicians, this issue becomes especially relevant due to the recent phenomenon of cross-border reproductive care – people cross borders for access to services that are cheaper, newer or legal compared to the ones in the country of origin (Martin, 2012).

The present study is the first attempting to establish comparability of infertility-related stress across genders and countries. Having a unique measure administrated in seven countries to samples of patients undergoing fertility treatment, we tested the following hypotheses: (i) the measure presents a structure with three factors, namely personal stress, marital stress, and social stress; (ii) these factors present good reliability and validity in all subsamples; (iii) the factor structure is invariant across genders and nationalities.

Method

Procedure

The COMPI research program (L. Schmidt, 2006; L Schmidt et al., 2003) was designed to evaluate the fertility treatment process and its psychosocial consequences among Danish fertility patients. It is the largest prospective cohort study ($n = 2250$) measuring infertile couples' motivations and expectations immediately before starting a treatment cycle.

At the time this retrospective study was being designed, COMPI instruments had been applied in 18 countries across America, Asia, and Europe by 23 researchers with consent from the original author. An initial and two follow-up requests were sent to these researchers asking for collaboration between March 2012 and August 2013, together with a brief online form. Of the 23 scholars contacted, nine research teams accepted to collaborate (see acknowledgments), one was still on the process of initiating data collection, five had never initiated data collection or had their projects

discontinued, six chose not to participate, and two never replied. The COMPI-FPSS was applied by seven research teams in the following countries: Denmark, Hungary, Croatia, China, Sweden Germany, and Greece. Data was collected in fertility centers between January 2000 and December 2013 (data collection for the Chinese dataset was still ongoing at the time permission was given and only sent in December). Each country's version of the COMPI-FPSS was translated from the English version provided by the original author and back-translated to English by an independent bilingual researcher. All versions were pre-tested with pilot studies or spoken reflection procedures.

Participants

After merging databases, the initial sample contained a total of 4171 subjects seeking fertility treatment in seven nations (see Table 1). There were more female ($n = 2370$, 57%) than male participants ($n = 1801$, 43%) due to exclusion of males in two countries, but also due to the usual higher attendance of women in fertility clinics. Participants were excluded if items left unanswered corresponded to $>10\%$ within a given dimension of the instrument or if they were extreme multivariate outliers ($|Mahalanobis D^2|, P. < .001$). After determining that data was missing at random within each country group ($P. > .05$), the remaining missing values were replaced by the sample's respective scale mean. No significant differences were observed between the imputed dataset and the original sample either for *personal stress*, *marital stress*, or *social stress* ($P. > .05$). The final sample included 3923 infertility patients (1691 men, 43% and 2232 women, 57%), with cultural subsamples ranging from 96 (China and Hungary) to 2068 (Denmark). Participants had a mean age of 33.9 years ($SD = 4.81$), with men being older ($M = 35.0$; $SD = 5.30$) than women ($M = 33.0$; $SD = 4.21$) ($t_{(3145,564)} = 12.530, p < .001$). The majority of participants (84.2%) were childless or

seeking treatment for primary infertility, and 15.8% already had children. Table 2 presents the mean, standard deviation, and standard error values for the FPSS domains by country and for the overall sample.

Measure

The *COMPI Fertility Problem Stress Scales* (L. Schmidt, 2006; L. Schmidt, Christensen, & Holstein, 2005) has 14 items measuring the amount of stress the fertility problem places on daily life. In addition to seven questions taken from the Fertility Problem Stress Inventory (Abbey et al., 1991), seven items were developed based on findings from The Psychosocial Infertility Interview Study (Schmidt, 1996). Then, an exploratory factor analysis produced a set of parsimonious factors and strain in relation to the three different domains (L Schmidt et al., 2003). The personal stress domain (six items) assesses the stress associated with infertility within the person's life and on mental and physical health; the marital stress domain (four items) measures the extent to which infertility produces strain on the marital and sexual relationships; and the social stress domain (four items) assesses the stress that infertility causes on social relations with family, friends, and workmates (see Supplementary Material). The response key for the personal and social domains and for two items from the marital domain is a four-point Likert scale from (1) "none at all" to (4) "a great deal". The response key for the remaining two items from marital stress was a five-point Likert response key from (1) "strongly disagree" to (5) "strongly agree". Higher scores indicate more stress.

Data analyses

Items were firstly computed into Z scores to allow direct comparison of items. Confirmatory factor analyses (CFA) were conducted for both country and gender groups to validate the a priori factor structure of the instrument, using AMOS 19 (IBM SPSS) and maximum likelihood estimation. Correlations between items were $r \leq .700$,

suggesting absence of multicollinearity. The goodness-of-fit was determined according to Hooper et al.'s (2008) goodness-of-fit cut-offs: $>.90$ for the comparative fit index (CFI), < 5 for the chi-square ratio (χ^2/df), and $< .07$ for the root mean square error of approximation (RMSEA). The modified expected cross-validation index (MECVI) was used to compare different models, with smaller values indicating better fit and stability for the population under study. Reliability and validity were analyzed following the guidelines of Hair and colleagues (2006). Reliabilities were determined by composite reliability (CR). CR uses not only the loadings but also the residuals of each item to verify the consistency of the items as manifestations of the latent variable, being therefore more appropriate for calculating reliability in SEM than Chronbach's alpha (Raykov, 1997). Convergent validity was determined by the average variance extracted (AVE) and by comparing CR to the AVE; discriminant validities were assessed by comparing the shared variance (squared correlation; SV) between each pair of factors against the AVEs of the two respective factors. Hair and colleagues (2006) provided the thresholds for these analyses: $CR > .70$ for reliability, $AVE > .50$ and $CR > AVE$ for convergent validity, and $AVE > SV$ for discriminant validity. Multi-group confirmatory factor analyses (MGCFA) were conducted to verify the structure invariance across country and gender, aiming to establish whether or not the COMPI-FPSS could be further used to compare individual scores among such groups. We repeated the analyses separately for country groups and gender groups, following the procedures laid out by Vandenberg and Lance (2000). After verifying whether or not the measurement model fitted the data well in all subgroups, equality constraints were imposed across the groups that we aimed to compare. The constraints were placed cumulatively on the measurement weights, the variances/covariances and the measurement residuals in order to test for differences among the models. Changes in the

models' CFI and RMSEA were used to determine invariance as recommended by Chen [38] ($\Delta\text{CFI} < -.010$, $\Delta\text{RMSEA} < .015$). Finally, critical ratios ($|Z| > Z_{0.975} = 1.96$) were used to verify pairwise differences between parameters. A critical ratio $> Z_{0.975}$ means that a given path (e.g., a certain factor loading) is significantly different between groups, and hence the respective item would be capturing the construct differently between groups. If full scalar invariance is not verified, partial scalar invariance must be established before groups can be compared (Byrne, 2013).

Results

Confirmatory factor analyses

Table 3 presents results concerning CFA models. First, we tested the non-specific, 14-item, three-factor original model (M1) separately in each country. The goodness-of-fit of this model was poor for all the country-samples. Five items were removed based on low factor loadings ($< .45$, as recommended by Tabachnick & Fidell, 2007), cross-factor loadings and multiple, high standardized residual covariances (see Supplementary material). Model 2 (M2) presented better fit in all subsamples. Modification indices suggested that freely estimating the residual covariance between items 2 and 3 (see Supplementary material) would yet improve fit in all subsamples, excepting for the Swedish and German subsamples. Given the conceptual similarity between items (both address the impact of the infertility experience in personal health), the residual correlation seemed justifiable, and therefore was maintained. The fit of this final model (M3; see Figure 1) to the data was excellent in Hungary and Greece and appropriate in Denmark, Sweden and Germany. However, models corresponding to Croatia and China did not yield an appropriate fit concerning the RMSEA index. Given

that no model respecifications improved the model to an acceptable fit in these cultures, participants from Croatia and China were excluded from further analyses.

Reliability, convergent validity and discriminant validity

Analysis concerning reliability and validity were conducted using only participants from countries whose models revealed acceptable to good fit indices, namely Denmark, Hungary, Germany and Greece (see Table 4). CR, AVE and SV were computed for each latent construct in the measurement model of the correspondent country. Reliability indices were good for the three factors in all countries, excepting for PS in Greece and MS in Sweden ($CR < .70$). The AVEs were above .50 for all subsamples, excepting for Sweden (in PS and MS) and Greece (in PS). However, all the AVEs were shorter than the respective CR, indicating convergent validity. Discriminant validity was verified for all factors in Denmark, Hungary and Germany as the AVEs of every factor were larger than the respective SV. The AVEs in Greece were above the variance shared by PS and MS, and the AVEs in Sweden were above the variance shared by PS and the other factors. Table 4 summarizes the reliability and validity results.

Invariance analyses

After verifying that the specified model fitted the data well in the subsamples, invariance analyses were performed across gender and country groups (Table 5). The model was found to be fully invariant across gender (configural, metric, and scalar invariance). Full invariance was not determined across countries. However, by freeing the loadings for “childlessness has caused crisis in our relationship” and for “how much stress has your fertility problem placed on your mental health?”, partial invariance was established ($|Z| > Z_{0.975} = 1.96$).

Discussion

This study attempted to validate a measure of stress specifically related to infertility- COMPI-FPSS. Previous suggestions that men and women, or individuals from different cultures, may differ in the way they experience infertility have compelled research to provide a validated fertility stress measure so that researchers and clinicians can be assured that the same construct is being assessed in each group. Our major goal was to establish COMPI-FPSS construct comparability across gender and cultures in order to allow comparison of these groups by future research. With a sample of men and women from seven different countries (Denmark, Hungary, Croatia, China, Sweden, Germany and Greece), we tested the structure, reliability, validity, and measurement invariance of the COMPI-FPSS and its three domains: personal stress, marital stress, and social stress.

The three-factor structure was confirmed for all subsamples. The final model contained three items per subscale, which were shown to present generally good convergent validity, discriminant validity and reliability. The items that were dropped might contain emotionally-evocative terms that might be particularly difficult to translate into different languages and interpreted by different cultures. However, the shortened version of the COMPI-FPSS (from 14 to 9 items) improved its parsimony and did not compromise its reliability. The AVEs larger than SVs in all the countries confirmed that the remaining items had more in common with the respective factor than they did with the remaining constructs, as expected. Furthermore, the AVEs indicated that variances in the items were essentially explained by the correspondent latent variable. The exceptions were for Greece and Sweden, in which variance in the items of personal stress and marital stress was found to be partially related to non-correspondent latent variables. Apparently, in Greece and Sweden, personal and marital stress can be

taken as fuzzy-bounded constructs, in the sense that the impact of stress on these individuals self-esteem might be more closely-related to the impact on marriage.

The respecified measurement model was psychometrically adequate and fitted the data well in all country-samples, with exception of Croatia and China. This poor Croatian and Chinese data fit can be due to methodological failures and/or important context differences among countries. Methodologically, we can have failed somehow in tapping all the control variables that may be relating to stress measurement while differing from the other country samples. For instance, Croatian and Chinese subsamples may have demographic characteristics that we did not control for and that influenced the way individuals interpreted the items. However, there also may be cultural, social, historical or political backgrounds justifying the observed cultural differences. For instance, the social memory of the political oppression climate and war that lasted up to the late 20th century and the consequent psychological sequelae that still persist (Brajša-Žganec, 2005) might justify the lack of invariance. Or, in China's case, restrictive birth policies there may be a cultural aspect that prevented Chinese from taking the construct of fertility-related stress similarly to individuals from countries where boosting birth rates is a goal to achieve.

After the decision to exclude Croatia and China from the remaining analyses, MGCFA demonstrated that the COMPI-FPSS was fully invariant across genders and partially invariant across the countries in study. In particular, we established configural, full metric and full scalar invariance across genders, and configural, partial metric and partial scalar invariance across the five countries. Although the measure did not meet criteria for residual invariance, many scholars suggest that equality on measurement residuals may be too conservative a constraint when analyzing measurement invariance (Dimitrov, 2010; Meredith & Teresi, 2006). Results pertaining to configural invariance

indicate that the construct of fertility-related stress as assessed in this study has a common meaning in the targets, so fertility-related stress can be assessed using these three three-item subscales in both genders and in all the five countries in study. Results on metric and scalar invariance demonstrated that seven of all nine items had equivalent factor loadings and that measurement models had similar structural variances/covariances across men and women (see table 5). Therefore, construct of infertility-stress and the items of the COMPI-FPSS seem to be understood by women and men similarly. Findings on non-invariance of Items 2 and 3 across countries (“Childlessness has caused crisis in our relationship” and “How much stress has your fertility problem placed on your mental health”) require further careful attention, as they seem to be interpreted differently in the different countries. Perhaps the non-invariance in such items about relationship crisis and mental health are due to the fact that different countries have different attitudes towards romantic relationships (Medora, Larson, Hortacsu, Hortagsu, & DAVE, 2002) and also conceptualize mental problems differently (Angst et al., 2010). Factors contributing specifically to these differences might be for example the fact that our dataset merged samples collected over ten years (conceptualizations of romantic relationships might evolve over a decade), or the lack of professional infertility counseling in some countries such as Denmark (conceptualizations of the impact on mental health might differ). Nevertheless, partial invariance is agreed to be sufficient in multinational cross-national studies like the present one (Steinmetz, 2011), as experts advocate that full measurement invariance across more than two culturally and linguistically different samples is scientifically unrealistic (De Beuckelaer & Swinnen, 2011; Steinmetz, 2011; Torsheim et al., 2012). In conclusion, our findings enable future studies to analyze differences either between genders or among these countries.

Although the COMPI-FPSS demonstrated promising results, our study reflected several limitations. First, some country-samples were relatively small (namely, Hungarian and Chinese), preventing us from stating firmer conclusions about them. Second, there is the possibility of significant differences in some variables that were not assessed in this study and that may be influencing the way individuals embedded in a specific culture interpret the items. A uniform definition of socio-economic status can be of value to any investigation with a cross-cultural focus and might have helped explain for example the drop of the Croatian and Chinese groups due to poor fit to the data. Third, this study included men and women seeking treatment, so the sample cannot be representative of infertile individuals who have never sought medical help. Fourth, after dropping the Chinese subsample, all countries in this study are European, so we cannot generalize our results to other world sites. Future studies looking at the psychometric properties of this instrument should try to assess predictive validity in order to compare the capability of these measures in detecting changes in fertility-specific adjustment over time, and in predicting important consequences such as depression after a failed cycle or post-partum depression. Further efforts will also be valuable in considering other variables that may be relevant for the experience of infertility and that this study failed in assessing besides socioeconomic status, such as, time attempting to conceive, infertility type and causation, stage of fertility treatment and early pregnancy loss.

Besides adding information on the reliability and validity of one of the most used psychosocial instruments to assess fertility patients, to our knowledge this is the first time a fertility stress instrument has been tested for measurement invariance across genders and cultures. As a validated, reliable, and short measure of fertility stress, the COMPI-FPSS can be regularly administered in both research and clinical settings to

target men and women suffering with infertility and its treatments in several cultural contexts. Being now adapted to both genders and five different countries, we believe the COMPI-FPSS to be able to generate new discussion on fertility-related stress.

Acknowledgments

The authors would like to thank the following researchers who very generously consigned their data for this study: Dr. Ginny He and Prof. Guangxiu. Lu, Reproductive and Genetic Hospital of Citic-Xiangya, China; Dr. Ivan Grbavac and Mariana Merkas, University Hospital Center Sestre Milosrdnice, Department of Gynecology and Obstetrics, Croatia; Jasmin Passet-Wittig, Ulrike Zier and Dr. Norbert Schneider, Federal Institute for Population Research, Germany; Dr. Claire Gourounti and Prof. Katerina Lykeridou, Technological Educational Institution of Athens , Department of Midwifery , Greece; Dr. Nikollet Papay, Budapest Catholic University , Department of Educational Science and Social Studies, Hungary; Carolyn Cesta, and Anastasia N. Iliadou, PhD, Karolinska Institutet, Institute of Medical Epidemiology & Biostatistics, Sweden; Selin Belen and Hande Sart, PhD., Boğaziçi University, Faculty of Psychology, Turkey.

Authors' roles

M.V.M. was responsible for data collection. M.P.S. was responsible for analysis and drafted the manuscript. All authors participated in the concept and design of the study, as well as interpretation of data, draft revisions and approval of manuscript submissions.

Funding

This study was supported by European Union Funds (FEDER/COMPETE – Operational Competitiveness Program), and by national funds (FCT – Portuguese Foundation for Science and Technology) under the projects PTDC/MHC-PSC/4195/2012 and SFRH/BPD/85789/2012.

Conflict of interests

None declared.

AUTHORS' COPY

References

- 1 Abbey, A., Andrews, F. M., & Halman, J. L. (1991). Gender's role in responses to infertility.
2 *Psychology of Women Quarterly*, 15(2), 295-316. doi:10.1111/j.1471-
3 6402.1991.tb00798.x
- 4 Angst, J., Meyer, T. D., Adolfsson, R., Skeppar, P., Carta, M., Benazzi, F., . . . YUAN, C. M. (2010).
5 Hypomania: a transcultural perspective. *World Psychiatry*, 9(1), 41-49.
- 6 Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process
7 of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191.
- 8 Bernstein, J., Potts, N., & Mattox, J. H. (1985). Assessment of psychological dysfunction
9 associated with infertility. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*,
10 14(s6), 63s-66s.
- 11 Brajša-Žganec, A. (2005). The long-term effects of war experiences on children's depression in
12 the Republic of Croatia. *Child Abuse and Neglect*, 29(1), 31-43.
- 13 Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic*
14 *concepts, applications, and programming*: Psychology Press.
- 15 Connolly, K. J., Edelmann, R. J., Cooke, I. D., & Robson, J. (1992). The impact of infertility on
16 psychological functioning. *Journal of Psychosomatic Research*, 36(5), 459-468.
- 17 De Beuckelaer, A., & Swinnen, G. (2011). Biased latent variable mean comparisons due to
18 measurement noninvariance: A simulation study. *European Association for*
19 *Methodology series European Association for Methodology series*, 117-147.
- 20 Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation.
21 *Measurement and Evaluation in Counseling and Development*, 43(2), 121-149.
- 22 Emery, M., Béran, M. D., Darwiche, J., Oppizzi, L., Joris, V., Capel, R., . . . Germond, M. (2003).
23 Results from a prospective, randomized, controlled study evaluating the acceptability
24 and effects of routine pre-IVF counselling. *Human Reproduction*, 18(12), 2647-2653.
25 doi:10.1093/humrep/deg501
- 26 Epstein, J., Santo, R. M., & Guillemin, F. (2015). A review of guidelines for cross-cultural
27 adaptation of questionnaires could not bring out a consensus. *Journal of Clinical*
28 *Epidemiology*, 68(4), 435-441.
- 29 Fontaine, J., & Fischer, R. (2010). Data analytic approaches for investigating isomorphism
30 between the individual-level and the cultural-level internal structure. *Cross-cultural*
31 *research methods in psychology*, 273-298.
- 32 Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data*
33 *analysis* (Vol. 6): Pearson Prentice Hall Upper Saddle River, NJ.
- 34 Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for
35 determining model fit. *Articles*, 2.
- 36 Kagami, M., Maruyama, T., Koizumi, T., Miyazaki, K., Nishikawa-Uchida, S., Oda, H., . . .
37 Schmidt, L. (2012). Psychological adjustment and psychosocial stress among Japanese
38 couples with a history of recurrent pregnancy loss. *Human Reproduction*, der441.
- 39 Lykeridou, K., Gourounti, K., Sarantaki, A., Loutradis, D., Vaslamatzis, G., & Deltsidou, A. (2011).
40 Occupational social class, coping responses and infertility-related stress of women
41 undergoing infertility treatment. *Journal of Clinical Nursing*, 20(13-14), 1971-1980.
- 42 Martin, L. (2012). Reproductive tourism. In G. Ritzer (Ed.), *Wiley-Blackwell Encyclopedia of*
43 *Globalization*.
- 44 Martins, M. V., Costa, P., Peterson, B. D., Costa, M. E., & Schmidt, L. (2014). Marital stability
45 and repartnering: infertility-related stress trajectories of unsuccessful fertility
46 treatment. *Fertility and Sterility*, 102(6), 1716-1722.

- 1 Martins, M. V., Peterson, B. D., Costa, P., Costa, M. E., Lund, R., & Schmidt, L. (2013).
2 Interactive effects of social support and disclosure on fertility-related stress. *Journal of*
3 *Social and Personal Relationships*, 30(4), 371-388. doi:10.1177/0265407512456672
- 4 Medora, N. P., Larson, J. H., Hortacsu, N., Hortagsu, N., & DAVE, P. (2002). Perceived attitudes
5 towards romanticism; a cross-cultural study of American, Asian-Indian, and Turkish
6 young adults. *Journal of Comparative Family Studies*, 155-178.
- 7 Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance.
8 *Medical Care*, 44(11), S69-S77.
- 9 Passet-Wittig, J., Letzel, S., Schneider, N. F., Schuhrke, B., Seufert, R., Zier, U., & Münster, E.
10 (2014). *The Pink Study – Methodology of the Baseline Survey of a Prospective Cohort*
11 *Study of Couples undergoing Fertility Treatment. BiB Daten und Methodenberichte*. .
12 Retrieved from
- 13 Peterson, B., Boivin, J., Norré, J., Smith, C., Thorn, P., & Wischmann, T. (2012). An introduction
14 to infertility counseling: a guide for mental health and medical professionals. *Journal of*
15 *Assisted Reproduction and Genetics*, 29(3), 243-248. doi:10.1007/s10815-011-9701-y
- 16 Peterson, B. D., Pirritano, M., Christensen, U., & Schmidt, L. (2008). The impact of partner
17 coping in couples experiencing infertility. *Human Reproduction*, 23(5), 1128-1137.
- 18 Pinborg, A., Hougaard, C., Andersen, A. N., Molbo, D., & Schmidt, L. (2009). Prospective
19 longitudinal cohort study on cumulative 5-year delivery and adoption rates among
20 1338 couples initiating infertility treatment. *Human Reproduction*, 24(4), 991-999.
- 21 Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied*
22 *Psychological Measurement*, 21(2), 173-184.
- 23 Sanchez, J. I., Spector, P. E., & Cooper, C. L. (2006). Frequently ignored methodological issues
24 in cross-cultural stress research *Handbook of multicultural perspectives on stress and*
25 *coping* (pp. 187-201): Springer.
- 26 Schmidt, L. (1996). *Psykosociale konsekvenser af infertilitet og behandling [Psychosocial*
27 *consequences of infertility and treatment]*. Copenhagen: FADL Press.
- 28 Schmidt, L. (2006). Infertility and assisted reproduction in Denmark: epidemiology and
29 psychosocial consequences. *Danish Medical Bulletin*, 53(4), 390-417.
- 30 Schmidt, L. (2009). Social and psychological consequences of infertility and assisted
31 reproduction-what are the research priorities? *Human Fertility*, 12(1), 14-20.
32 doi:10.1080/14647270802331487
- 33 Schmidt, L., Christensen, U., & Holstein, B. E. (2005). The social epidemiology of coping with
34 infertility. *Human Reproduction*, 20(4), 1044-1052.
- 35 Schmidt, L., Holstein, B., Boivin, J., Sångren, H., Tjørnhøj-Thomsen, T., Blaabjerg, J., . . .
36 Rasmussen, P. (2003). Patients' attitudes to medical and psychosocial aspects of care
37 in fertility clinics: findings from the Copenhagen Multi-centre Psychosocial Infertility
38 (COMPI) Research Programme. *Human Reproduction*, 18(3), 628-637.
- 39 Schmidt, L., Holstein, B. E., Christensen, U., & Boivin, J. (2005). Communication and coping as
40 predictors of fertility problem stress: cohort study of 816 participants who did not
41 achieve a delivery after 12 months of fertility treatment. *Human Reproduction*, 20(11),
42 3248-3256. doi:10.1093/humrep/dei193
- 43 Schmidt, L., Holstein, B. E., Christensen, U., & Boivin, J. (2005). Does infertility cause marital
44 benefit?: An epidemiological study of 2250 women and men in fertility treatment.
45 *Patient Education and Counseling*, 59(3), 244-251.
- 46 Steinmetz, H. (2011). Estimation and comparison of latent means across cultures. *Cross-*
47 *cultural analysis: Methods and applications*, 85-116.
- 48 Torsheim, T., Samdal, O., Rasmussen, M., Freeman, J., Griebler, R., & Dür, W. (2012). Cross-
49 national measurement invariance of the teacher and classmate support scale. *Social*
50 *indicators research*, 105(1), 145-160.

- 1 Uysal-Bozkir, Ö., Parlevliet, J. L., & de Rooij, S. E. (2013). Insufficient cross-cultural adaptations
2 and psychometric properties for many translated health assessment scales: A
3 systematic review. *Journal of Clinical Epidemiology*, 66(6), 608-618.
- 4 van Empel, I. W. H., Nelen, W. L. D. M., Tepe, E. T., van Laarhoven, E. A. P., Verhaak, C. M., &
5 Kremer, J. A. M. (2010). Weaknesses, strengths and needs in fertility care according to
6 patients. *Human Reproduction*, 25(1), 142-149. doi:10.1093/humrep/dep362
- 7 Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance
8 literature: Suggestions, practices, and recommendations for organizational research.
9 *Organizational research methods*, 3(1), 4-70.

AUTHORS' COPY

1

2 Table 1

3 *Socio-demographic characteristics of the sample by country.*

Country	Language	<i>N</i>	% of females	Age (y)	Parity
				Mean (SD)	
Denmark	Danish	2068	52.2%	33.1 (4.48)	15.3%
Hungary	Hungarian	96	69.8%	34.8 (4.90)	10.4%
Croatia	Croatian	199	50.3%	35.2 (4.96)	9.5%
China	Chinese	96	100.0%	33.1 (4.68)	7.3%
Sweden	Swedish	778	55.4%	34.7 (4.84)	20.6%
Germany	German	531	57.3%	34.2 (5.34)	14.9%
Greece	Greek	155	100.0%	37.0 (4.22)	9.0%
Total	-	3923	56.9%	33.9 (4.81)	15.4%

4

5

1 Table 2
 2 *Mean, standard deviation, and standard error values for the FPSS domains by country*
 3 *using maximum likelihood estimation.*

4

Country	Personal Stress			Marital Stress			Social Stress		
	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>
Denmark	2.125	0.790	0.017	2.205	0.922	0.020	1.507	0.658	0.014
Hungary	2.118	0.752	0.077	1.455	0.545	0.056	1.365	0.441	0.045
Greece	2.611	0.830	0.067	1.871	0.724	0.058	1.525	0.604	0.049
Croatia	1.851	0.658	0.047	1.441	0.527	0.037	1.260	0.474	0.034
China	1.955	0.723	0.074	2.653	0.862	0.088	2.688	0.842	0.086
Sweden	1.948	0.611	0.022	1.632	0.490	0.018	1.309	0.381	0.014
Germany	2.494	0.913	0.040	2.289	0.795	0.035	1.630	0.681	0.030
Total	2.141	0.793	0.013	2.044	0.856	0.014	1.498	0.644	0.010

5

Table 3

Fit indexes for confirmatory factor analyses of the 9-item version of the COMPI-FPSS in samples from the countries in study.

Model	Sample																				
	Denmark			Hungary			Greece			Croatia			China			Sweden			Germany		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
fit index																					
<i>df</i>	74	24	23	74	24	23	74	24	23	74	24	23	74	24	23	74	24	23	74	24	23
χ^2/df	24.777	6.600	5.424	2.497	1.162	1.033	3.296	1.597	1.339	6.279	2.966	2.660	3.371	1.874	1.950	11.149	5.010	5.121	8.927	3.719	3.793
CFI	.845	.978	.983	.776	.981	.996	.705	.949	.972	.749	.941	.953	.736	.927	.924	.771	.938	.936	.821	.961	.961
RMSEA	.107	.052	.047	.126	.041	.019	.122	.049	.047	.163	.100	.092	.158	.092	.100	.114	.072	.073	.122	.072	.073
MECVI	0.917	0.097	0.083	2.720	0.788	0.768	2.030	0.540	0.506	2.686	0.583	0.543	3.401	0.968	0.990	1.143	0.206	0.209	1.367	0.249	0.249

Note. M1 = Model 1 = original 14-item, three-factor model; M2 = Model 2 = specified nine-item, three-factor model; M3 = Model 3 =

respecified nine-item, three-factor model with a modification based on sample-specific models (i.e., residual covariance of Item 2 with Item 3).

CFI, comparative fit index; RMSEA, root mean square error of approximation; MECVI, modified expected cross-validation index.

Table 4

Composite reliabilities (CR), average variance extracted (AVE) and shared variance (SV) for the COMPI-FPSS by country.

Country	Factor								
	Personal Stress (PS)		Marital Stress (MS)		Social Stress (SS)		Factors' SV		
	CR	AVE	CR	AVE	CR	AVE	SV _{PS,MS}	SV _{PS,SS}	SV _{MS,SS}
Denmark	.933	.518	.845	.652	.883	.717	.452	.338	.199
Hungary	.739	.493	.761	.517	.846	.652	.410	.132	.211
Sweden	.700	.441	.629	.384	.859	.674	.326	.465	.252
Germany	.851	.679	.790	.562	.861	.677	.490	.312	.376
Greece	.652	.399	.733	.491	.820	.607	.728	.220	.086

Table 5

Multigroup confirmatory factor analyses of the COMPI-FPSS's invariance by country and gender (Denmark, Hungary, Sweden, Germany and Greece).

Test	Invariance	Model fit					Invariance test	
		χ^2	df	χ^2/df	RMSEA	CFI	ΔCFI	$\Delta RMSEA$
Invariance by country	Configural	545.994	136	4.015	.029	.945	-	-
	Metric	784.336	154	5.093	.034	.921	-.024	.005
	Scalar	1444.223	172	8.397	.045	.854	-.067	.011
	Residual	3123.475	203	15.387	.063	.683	-.171	.018
Invariance by gender	Configural	338.940	46	7.368	.040	.974	-	-
	Metric	402.090	52	7.732	.041	.968	-.006	.001
	Scalar	494.350	58	8.523	.044	.961	-.007	.003
	Residual	1258.210	68	18.503	.029	.893	-.068	.015

CFI, comparative fit index; RMSEA, root mean square error of approximation

CFI, comparative fit index; RMSEA, root mean square error of approximation

Suppl. material: Final items and response keys for the COMPI-FPSS.*Personal domain*

2. It is very stressful for me to deal with this fertility problem.

How much stress has your fertility problem placed on the following:

5. your physical health?

6. your mental health?

Marital domain

What consequences has your childlessness for your marriage/partnership?

7. Childlessness has caused a crisis in our relationship.

How much stress has your fertility problem placed on the following:

9. your marriage/partnership?

10. your sex life?

Social domain

How much stress has your fertility problem placed on the following:

11. your relationships with your family?

12. your relationships with your family in law?

13. your relationships with friends?

Response key for items 2 and 7: (1) strongly disagree, (2) somewhat disagree, (3) neither agree nor disagree, (4) somewhat agree, (5) strongly agree.

Response key for items 5, 6, 9-13: 1) not at all, (2) a little, (3) some, (4) a great deal.

Note. Removed items:

Personal domain

1. My life has been disrupted because of this fertility problem

How much stress has your fertility problem placed on the following:

3. Your relationships with people with children?

4. Your relationships to pregnant women?

Marital domain

What are the consequences of childlessness for your marriage/relationship?

8. Childlessness has given thoughts about divorce.

Social domain

How much stress has your fertility problem placed on the following:

14. Your relationships with workmates?

Suppl. material Figure 1

Representations of Model 1 and Model 3.

