Journal of Articles in Support of the Null Hypothesis
Vol. 13, No. 2

www.jasnh.com

An analysis of measurement invariance in work stress by sex: Are we comparing apples to apples?

Amber Bielecky[1]
Selahadin Ibrahim[1,2]
Cameron A. Mustard[1,2]
Chantal Brisson[3,4]
Peter M. Smith[1,2,5]

[1] Institute for Work and Health, Toronto, Canada
[2] Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
[3] CHU de Québec Research Centre (CRCHU), Quebec City, Canada
[4] Department of Social and Preventive Medicine, Faculty of Medicine, Laval University, Quebec City, Canada
[5] School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

This study explores whether instruments used to measure *Job Control* and *Psychological Job Demands* function differently by sex. We evaluated the comparability of these measures by sex using multigroup confirmatory factor analysis. All estimates were invariant by sex, except for the intercept for the item assessing how "hectic" a worker perceives his/her job (which contributes to the Demands measure). This item intercept was statistically significantly higher for women than men, but not practically so. These findings suggest that the differences observed between men and women in the impact of work stress on health cannot be explained by measurement bias (when using the work stress measures employed in this study). Further research is needed to understand why work stress appears to affect the health of men and women differently.

Keywords: Psychosocial Work Stress; Multigroup Confirmatory Factor Analysis; Sex; Measurement; Psychological Demands; Job Control

Address of Corresponding Author:
Peter Smith (psmith@iwh.on.ca)
481 University Ave., Suite 800, Toronto, ON, Canada, M5G 2E9
T – 416-927-2027(x2226), F – 416-927-4167

Introduction

The impact of the psychosocial work environment on health has been the subject of scientific investigations for several decades (Eller et al., 2009; Gilbert-Ouimet, Trudel, Brisson, Milot, & Vezina, 2014; Karasek & Theorell, 1990; Kivimaki et al., 2006; van der Doef & Maes, 1999). In more recent years, attention has turned to how this environment may impact men and women differently (Cesana et al., 2003; Heraclides, Chandola, Witte, & Brunner, 2009; Heraclides, Chandola, Witte, & Brunner, 2012; Landsbergis, Dobson, Koutsouras, & Schnall, 2013; Norberg et al., 2007; Ohlin, Berglund, Rosvall, & Nilsson, 2007; Smith, Glazier, Lu, & Mustard, 2012; Smith, Mustard, Lu, & Glazier, 2013; Trudel, Brisson, & Milot, 2010). Some studies have found that the negative impact of low job control and high job strain on cardiovascular outcomes (such as hypertension) may be greater for men than women (Cesana et al., 2003; Landsbergis et al., 2013; Ohlin et al., 2007; Smith et al., 2013; Trudel et al., 2010). In contrast, low job control may increase women's risk for diabetes, but not men's (Heraclides et al., 2009, 2012; Norberg et al., 2007; Smith et al., 2012). One potential explanation for this varied effect may be that the instruments used to measure the psychosocial work environment function systematically differently for men versus women. If present, such a difference would reduce the validity of any sex-based comparisons about the impact of the psychosocial work environment (Drasgow & Kanfer, 1985). If the instruments are shown to be equivalent for men and women, then research can confidently focus attention on understanding the true source of the differential impact of work stress. This understanding would, in turn, inform prevention efforts, perhaps suggesting sex-specific activities. However, before spending valuable research resources hunting down the source of the differential impact of work stress on health, it is prudent to generate evidence in support of the null hypothesis, which would assume no sex-based differences in the properties of the work stress measures.

Since the psychosocial work environment is not directly observable, multi-item instruments are used to capture the characteristics thought to represent the underlying unobserved constructs, also known as "factors." If a measure functions as expected, each individual item in the instrument is thought to be related to a single underlying factor. While an individual item is an imperfect measure of the specific factor it relates to, the whole set of items together are thought to provide a valid representation of that factor. When this is the case, the individual items can be summed to form composite scores (one for each factor), and these composite scores accurately represent variation in the underlying factors among respondents (Nunnally, 1978). However, it is possible that the relationships between each item and the factors differ by sex (Meredith & Teresi, 2006). If such differences exist, the same composite work stress score would actually represent different levels of true underlying work stress for men versus women or, put another way, the scale of composite work stress score would be different for men and women. There are four facets of the relationships (between items and factors), and equivalence in each facet corresponds to a different type of measurement invariance (Meredith & Teresi, 2006). Table 1 lists these four types of measurement invariance, with the last column indicating how each type of invariance would appear in the relationship between sex and work stress measurement.

Table 1. *Types of Measurement Invariance*

Type of Measurement Invariance	Facet of the Relationship Between Items and Factors	How Does This Type of Invariance Appear in the Relationship Between Sex and Work Stress Measurement?
Configural invariance	Factor patterns = how many factors exist and which items are related to which factors	The number of work stress factors that exist, and the ways the work stress items are related to the work stress factors, are the same for men and women.
Metric invariance	Factor loadings = the strength of the relationship between each item and its corresponding factor	The strength of the relationship between each work stress item and its corresponding work stress factor is equal for men and women
Scalar invariance	Item intercepts = the mean value of items when the underlying factors are equal to zero	The mean values of the work stress items are the same for men and women when the underlying work stress factors are equal to zero.
Strict invariance	Error terms = the variation remaining after taking into account the relationship between items and factors	The error terms (which represent the variation remaining after taking into account the relationships between work stress items and corresponding work stress factors) are the same for men and women.

The Job Contents Questionnaire (JCQ) is one of the most widely used instruments to measure the psychosocial work environment (Karasek et al., 1998; Sanne, Torp, Mykletun, & Dahl, 2005). The two main factors arising from this questionnaire are decision latitude (also referred to as *Job Control*) and psychological demands (also referred to as *Demands*); the decision latitude factor consists of two lower-order factors (decision authority and skill discretion), but most research on psychosocial work stress focuses on the higher-order factor of decision latitude (or *Job Control*). The full version of the questionnaire uses nine items to measure *Job Control* and another nine to measure *Demands*; an older version of the JCQ employed only five items to measure *Demands* (Karasek et al., 1998).

Researchers in Canada used the JCQ to inform the development of an extremely brief instrument to measure the psychosocial work environment, which is used in the main population-based household surveys focused on health in Canada (Statistics Canada, 2014). In this instrument, five items are used to measure *Job Control* and two items are used to tap in to *Demands* (the exact wording of the items are shown in Table 2). In addition to a difference in the number of items used to measure each factor, this brief instrument differs from the original JCQ in that the Canadian instrument provides five response categories (0-*Strongly Agree*, 1-*Agree*, 2-*Neither Agree nor Disagree*, 3-*Disagree*, 4-*Strongly Disagree*) instead of the four used in the original (which are worded the same, but do not include the *Neither Agree nor Disagree* option). The *Job Control* items are then summed to create a composite *Job Control* score (which has a theoretical range of 0 to 20) and, separately, the items used to tap in to *Demands* are summed to create a composite *Demands* score (which has a theoretical range of 0 to 8) (Statistics Canada, 2014).

Two authors have alluded to variance by sex in the JCQ (Gaudette, 1998; Karasek et al., 1998). Specifically, Karasek et al. (1998) reported that some of the items used to measure Demands did not load clearly for men in a Canadian dataset that used the full nine-item

Table 2. *Items in the Psychosocial Work Environment Measure Used in Main Canadian Health Surveys*

Factor	Item *(Short Name)*
Job Control	1. Your job required that you learn new things ***(Learn)*** 2. Your job required a high level of skill ***(Skill)*** 3. Your job allowed you freedom to decide how you did your job ***(Freedom)*** 4. Your job required that you do things over and over ***(Repetitive)*** 5. You had a lot to say about what happened in your job ***(Say)*** *NOTE: Before calculating a job control factor score, items 1, 2, 3, and 5 are reverse scored, so that higher scores indicate more job control, and lower scores less job control; this reverse coding is also reflected in covariance terms examined in the multigroup confirmatory factor analysis conducted in this study.*
Demands	1. Your job was very hectic ***(Hectic)*** 2. You were free from conflicting demands that others made ***(Conflict)*** *NOTE: Before calculating a demands factor score, item 1 is reverse scored, so that higher scores indicate more demands and lower scores less demands; this reverse coding is also reflected in covariance terms examined in the multigroup confirmatory factor analysis conducted in this study.*

measure, and Gaudette (1998) stated that the two items used to measure Demands (in the brief Canadian measure) related to the underlying factors differently for men and women. However, both of these analyses relied on stratified Exploratory Factor Analysis (EFA), which is not intended for assessing measurement invariance across groups. The appropriate method for such assessments is a multigroup confirmatory factor analysis (MGCFA) because it provides a statistical test for *each* type of measurement invariance (Joreskog, 1971; Wu, Li, & Zumbo, 2007), whereas stratified EFA simply suggests the potential for invariance and, even then, only of two types (configural and metric). We were unable to find any studies that used a MGCFA to test for measurement invariance by sex using the full or abbreviated versions of the JCQ. We aim to fill this gap by conducting a MGCFA on data from the 2012 Canadian Community Health Survey to test for measurement invariance by sex in the brief Canadian measures of *Job Control* and *Demands*.

Method

Sample

The study uses cross-sectional data arising from the 2012 Canadian Community Health Survey (CCHS) (Statistics Canada, 2013). The 2012 CCHS used a multi-staged, stratified sampling frame to target individuals aged 12 and over, living in private dwellings within Canada. People living on Indian reserves or Crown lands, residents of institutions, full-time members of the Canadian Armed Forces and residents of certain remote regions were excluded from the sampling frame. The household response rate was 79.8%, and the selected-person response rate was 86.3%, producing a sample of 25,113 respondents.

For the purpose of the analyses presented in this paper, we restricted our sample to those respondents who were 20-74 years of age and were employed. To create a more homogeneous sample with respect to psychosocial work environment exposures, we also restricted our sample to those who were: (1) not self-employed, (2) usually worked more than 15 hours per week, and (3) not currently attending school. As a result, 9,495 respondents were eligible for the analysis. After excluding those respondents missing responses to the psychosocial work environment questions (n = 184, or 2% of the eligible respondents), 9,311 respondents were available for the analyses presented in this paper (4,569 men and 4,742 women). Approval for the secondary data analyses was obtained through the University of Toronto, Health Sciences I Ethics Committee.

Measure of the Psychosocial Work Environment

As noted in the Introduction, the CCHS used five items to measure *Job Control* and two to measure *Demands* (see Table 2).

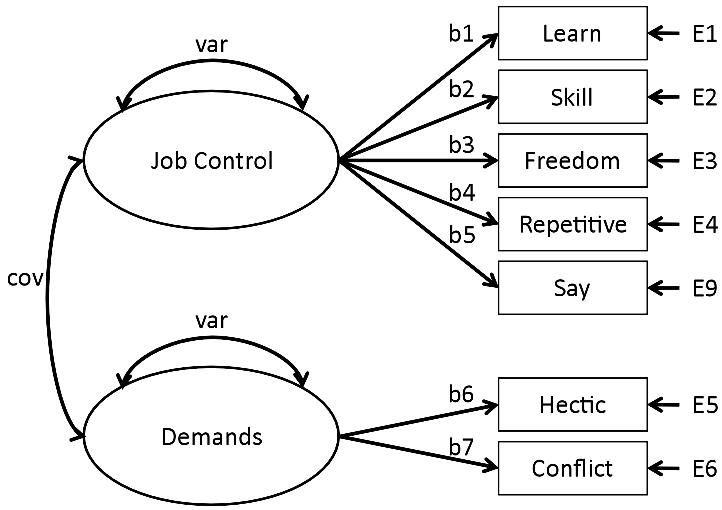
Statistical analyses

Multigroup confirmatory factor analysis (MGCFA) is the most widely used statistical technique for assessing measurement invariance (Joreskog, 1971; Wu et al., 2007). This technique is carried out in two sequential phases. In the first phase, a baseline model is built for each of the groups to be compared and, in the second phase, the various types of measurement invariance are tested in a hierarchical fashion (as listed in Table 1), with restrictions added when testing one type of invariance retained when testing all subsequent types. We conducted MGCFA using Mplus, employing the robust maximum likelihood (MLR) method of estimation to accommodate the non-normality in the item response data (Muthen & Muthen, 2010). Although (as noted above) the items used to measure *Job Control* and *Demands* use a 5-point Likert response scale and are, therefore, not technically continuous, previous work has shown that MLR estimation may be used reliably with this type of response data (Rhemtulla, Brosseau-Liard, & Savalei, 2012). All analyses were weighted to account for the probability of selection into the sample and initial survey non-response (Statistics Canada, 2013). Unfortunately, it was not possible to adjust the standard errors to account for the complex sample design because Mplus does not allow for bootstrapping when using the MLR method of estimation. As a result, the standard errors reported below are somewhat underestimated.

For our analyses, we built sex-specific baseline models by estimating the theoretical measurement model (see Figure 1) and examining four goodness-of-fit indices (the Root Mean Square Error of Approximation (RMSEA); the Standardized Root Mean Square Residual (SRMR); the Comparative Fit Index (CFI); and the Tucker-Lewis Index (TLI)) (Byrne, 2012). We then adjusted the sex-specific models until we achieved values indicative of good model fit: RMSEA and SRMR $\leq$ 0.05, and CFI and TLI $\geq$ 0.95 (Hu & Bentler, 1999). Since our goal was to assess measurement invariance in a model that reflected how the data were used in practice, we did not permit cross-loadings (where an item loads on to a factor other than the one it is theorized to load on) when adjusting the sex-specific models. We did allow for correlations between the item error terms (also referred to as residual covariance terms) in order to improve the goodness-of-fit indices. As baseline models for men and women were built separately, sex-specific models may include different residual covariance terms.

We then tested the different types of measurement invariance in a hierarchical manner, as outlined in Table 1, starting with configural invariance. After pooling the data

Figure 1: *Hypothesized Measurement Model*



for men and women, we estimated a single model that included the theorized measurement model along with the sex-specific residual covariance terms (added when building the sex-specific baseline models). To assess configural invariance, the factor patterns were constrained to be equal across men and women, while the factor loadings, item intercepts, item variance terms and residual covariance terms, were free to vary by sex. The four fit indices described above provided the formal test of configural invariance, with support for invariance provided when these indices suggested a good fit (as noted above, when RMSEA and SRMR ≤ 0.05, and CFI and TLI ≥ 0.95). To test for metric, scalar and strict invariance, we used the partial measurement invariance approach described by Byrne, Shavelson, and Muthen (1989) and the change (Δ) in CFI test described by Cheung and Rensvold (2002).

The partial measurement invariance approach involves testing for each type of measurement invariance, while allowing for specific differences across groups (Byrne et al., 1989). For example, if all but one factor loading is equal across groups, the partial measurement invariance approach would allow that single factor loading to vary across groups, constrain the remaining factor loadings to be equal, and continue to assess the remaining types of measurement invariance (scalar and strict). The Δ in CFI test is recommended for studies involving large samples, and Δ values of less than 0.010 provide support for invariance (Cheung & Rensvold, 2002). To describe how partial measurement invariance and the Δ in CFI test work, the following sentences walk through the steps we used to assess metric invariance. First, we estimated a model where the factor loadings were constrained to be equal across men and women. Then, we calculated the Δ in the CFI between this model and the model used to assess configural invariance. If the Δ in the CFI was equal to or greater than 0.010, we determined which factor loading was different (using the Modification Indices produced by Mplus) and re-generated the model, allowing this factor loading to vary by sex. We repeated this process until achieving a Δ in the CFI of less than 0.010 (when compared to the model used to assess configural invariance). The same approach was used for assessing scalar and strict invariance, except the comparison model for each was different; the final model fit when assessing metric invariance was the comparator for scalar invariance tests, while the final model fit when assessing scalar invariance was the comparator for strict invariance tests.

Results

Planned analyses

When building the sex-specific baseline models, it was necessary to add six residual covariance terms for men, and five for women, to achieve good fit with the sex-specific data; all residual covariance terms added were deemed theoretically plausible by the research team. Three of these terms were common between men and women [(i) *Freedom* with *Say*, (ii) *Repetitive* with *Hectic*, and (iii) *Freedom* with *Conflicting*], three were unique to men [(i) *Hectic* with *Say*, (ii) *Repetitive* with *Say*, and (iii) *Freedom* with *Repetitive*) and two were unique to women ((i) *Conflicting* with *Say* and (ii) *Freedom* with *Hectic*). It was necessary to fix the residual variance term for one item (namely, for the *Hectic* item) in the model for men. This fixing was necessary because the relationship between this item and its corresponding factor (*Demands*) was so strong that it accounted for more than 100% of the variance and, as a result, the residual variance term was negative; this type of relationship is referred to as a Heywood case. Fixing the variance of Heywood cases is acceptable when the confidence interval for the residual variance term includes zero (Schumacker & Lomax, 2010), which was the case in our results.

Table 3 presents the results for the tests of each type of measurement invariance by sex. The values of the four model fit indices for the initial model provide support for configural invariance by sex, with RMSEA = 0.023, SRMR = 0.025, CFI = 0.979 and TLI = 0.949. Upon constraining the factor loadings, the Δ in the CFI value was less than 0.010, which provides support for metric invariance. Upon constraining the intercepts, the Δ in the CFI value was greater than 0.010, indicating that not all intercepts were equal. The Modification Indices suggested the intercept for the Hectic item was statistically different between men and women. Upon allowing that intercept to vary by sex, while keeping all other intercepts equal, the Δ in the CFI value was less than 0.010, which provides support for scalar invariance for all remaining item intercepts. Finally, upon constraining all residual variance and common covariance terms, the Δ in the CFI value was less than 0.010, which provides support for strict invariance.

Table 3: *Tests of Measurement Invariance By Sex*

Model #	Type of Measurement Invariance Assessed	Modification	CFI	Δ in CFI	Comparator Model #
1	Configural Invariance	---	0.979	---	---
2	Metric Invariance	Factor loadings set equal by sex	0.977	-0.002	1
3	Scalar Invariance	Intercepts set equal by sex	0.950	-0.027	2
4	Scalar Invariance	*Hectic* intercept free to vary by sex	0.968	-0.009	2
5	Strict Invariance	Residual and residual covariance terms set equal by sex	0.970	+0.002	4

Note: Comparative Fit Index (CFI)

Table 4 presents the unstandardized factor loading, intercept, residual variance and residual covariance estimates from the final model, along with the corresponding standard errors and p-values. As noted above, all estimates were invariant by sex, except for the *Hectic* item intercept, which contributes to the *Demands* factor. This intercept was 0.244 points (=3.819 − 3.575) higher for women compared to men. This value represents 3% of the total range of scores possible for the composite *Demands* factor score (which has a theoretical range of 0 to 8), and means that the same observed *Demands* score represents slightly higher true underlying job demands for women compared to men.

Post-hoc analyses

To confirm that fixing the variance for the *Hectic* item did not cause the statistical difference observed in the intercept for that item, we generated sex-specific estimates using Bayesian methods. The Bayesian sex-specific estimates were comparable to the MLR multigroup estimates in which the variance was fixed for the *Hectic* item (results available upon request).

Discussion

This study examined the brief measures of *Job Control* and *Demands* used in the main population-based household surveys focused on health in Canada. While all factor loadings, residual variance terms and residual covariance terms were equal between men and women, one item intercept differed by sex. The sex-variant intercept was for the *Hectic* item, which contributes to the measure of *Demands*. This intercept was 0.244 points (3% of the *Demands* score range) higher for women compared to men, and means that the same observed Demands score represents slightly higher true underlying job demands for women compared to men. These findings suggest that sex-based comparisons based on the composite *Job Control* score are valid, but that sex-based comparisons involving the composite *Demands* score may be biased. The anticipated impact of this potential bias is discussed below.

As noted in the introduction, because the psychosocial work environment is not directly observable, multi-item instruments are used to capture "factors" thought to represent this environment. Each individual item in such an instrument is thought to be related to one of the factors. The relationship between each item and its corresponding factor is reflected in a regression equation. In this equation, the item is the outcome, the factor is the main independent variable and the intercept is the value of the item when the factor is equal to zero. Our results indicated that the intercept in the equation relating the *Hectic* item to the *Demands* factor was different for men and women. To help conceptualize how this difference would impact sex-based comparisons of the composite *Demands* score, we draw upon an example used by Gregorich (2006) focused on patient weight. Imagine a clinic where the procedure used for measuring weight depends on patient sex: while female patients in the clinic are asked to remain in their street clothes when weighed (with street clothes weighing an average of 2.2lbs), men are asked to don an examination gown before being weighed (with such gowns weighing an average of 0.2lbs). These sex-specific procedures introduce a systematic bias (of, on average, 2lbs) into the observed different in weights between men and women. Returning to our finding for the *Hectic* intercept, we would anticipate sex-based comparisons of the *Demands* score to be biased by a difference

Table 4: *Unstandardized Factor Analysis Estimates for Men and Women*

Type	Specific	Applies to men and women	Estimate	SE	p-value
Factor Loading	*Learn* on Job Control	yes	0.555	0.024	<0.001
	Skill on Job Control	yes	0.821	0.026	<0.001
	Freedom on Job Control	yes	0.351	0.023	<0.001
	Repetitive on Job Control	yes	0.117	0.023	<0.001
	Say on Job Control	yes	0.384	0.023	<0.001
	Hectic on Demands	yes	1.065	0.012	<0.001
	Conflicting on Demands	yes	0.220	0.023	<0.001
Item Intercept	*Learn*	yes	4.216	0.019	<0.001
	Skill	yes	4.153	0.017	<0.001
	Freedom	yes	3.830	0.018	<0.001
	Repetitive	yes	2.237	0.019	<0.001
	Say	yes	3.657	0.019	<0.001
	Hectic	Men ONLY	3.575	0.027	<0.001
	Hectic	Women ONLY	3.819	0.028	<0.001
	Conflicting	yes	2.894	0.022	<0.001
Residual Variance	*Learn*	yes	0.580	0.035	<0.001
	Skill	yes	0.206	0.037	<0.001
	Freedom	yes	0.925	0.027	<0.001
	Repetitive	yes	1.141	0.026	<0.001
	Say	yes	0.949	0.025	<0.001
	Hectic	yes	0.063	0.000	---[a]
	Conflicting	yes	1.340	0.024	<0.001
Residual Covariance	*Freedom* with *Say*	yes	0.333	0.020	<0.001
	Freedom with *Conflicting*	Yes	-0.181	0.021	<0.001
	Repetitive with *Hectic*	Yes	-0.190	0.022	<0.001
	Hectic with *Say*	Men ONLY	0.126	0.025	<0.001
	Repetitive with *Say*	Men ONLY	0.121	0.026	<0.001
	Freedom with *Repetitive*	Men ONLY	0.088	0.029	<0.001
	Conflicting with *Say*	Women ONLY	-0.160	0.030	<0.001
	Freedom with *Hectic*	Women ONLY	-0.115	0.023	<0.001

[a] *There is no-p-value associated with the residual variance term for the Hectic item because, as explained in the text, this term was fixed to allow for model estimation in the presence of a Heywood case.*

of 0.244, or 3% of the total range possible for the *Demands* score. Although *statistically* significant, a potential bias of this magnitude is unlikely to be of *practical* significance.

Given the only difference observed between men and women was negligible, this study has generated evidence in support of the null hypothesis: there are no sex-based differences in the properties of the work stress measures examined in this study. By eliminating this competing hypothesis, this evidence provides the necessary foundation for studies aimed at understanding the true source of the differential impact of work stress. Studies focused on other explanations for the sex differences will, in turn, generate evidence to inform sex-specific prevention efforts.

However, it is important to note that this study only examined the two-item *Demands* measure, and previous studies have shown that this reduced-item measure was of limited validity relative to the scores generated using the full (nine-item) or fuller (five-item) versions of the *Demands* scale factor (Brisson & Larocque, 2001; Gaudette, 1998). It is possible that using the full (or fuller) version of the *Demands* scale would produce different sex-based results than those observed in this study.

Conclusion

The measure of *Job Control* used in the 2012 CCHS is invariant by sex. Although there is evidence of *statistically* significant variation by sex in the measure of *Demands*, this variation is unlikely to be *practically* significant. Therefore, comparisons drawn between men and women based on either the *Job Control* or *Demands* measures used in the 2012 CCHS are valid. These findings suggest that the differences observed between men and women in the impact of work stress on health cannot be explained by measurement bias, though an analysis using the full *Demands* scale is warranted. Further research is needed to understand why work stress appears to affect the health of men and women differently.

References

Brisson, C., & Larocque, B. (2001). Validité des indices de demande psychologique et de latitude décisionnelle utilisés dans l'Enquête nationale sur la santé de la population (ENSP) de 1994-95 [Validity of the application indexes psychological and latitude used in decision National Survey on Health population (NPHS) 1994-1995]. *Canadian Journal of Public Health, 92*(6), 468-474.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Multivariate applications series. New York: Taylor & Francis Group.

Byrne, B. M., Shavelson R.J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466.

Cesana, G., Sega, R., Ferrario, M., Chiodini, P., Corrao, G., & Mancia, G. (2003). Job strain and blood pressure in employed men and women: A pooled analysis of four Northern Italian population samples. *Psychosomatic Medicine, 65*, 558-563.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological-measurement in heterogeneous populations. *Journal of Applied Psychology, 70*(4), 662-680.

Eller, N. H., Netterstrom, B., Gyntelberg, F., Kristensen, T. S., Nielsen, F., Steptoe, A., & Theorell, T. (2009). Work-related psychosocial factors and the development of ischemic heart disease. A systematic review. *Cardiology in Review, 17*(2), 83-97.

Gaudette, L. (1998). *Validity and reliability of a 12-item version of the Karasek job characteristic scale used in the National Population Health Survey 1994/95*. Ottawa: Statistics Canada.

Gilbert-Ouimet, M., Trudel, X., Brisson, C., Milot, A., & Vezina, M. (2014). Adverse effects of psychosocial work factors on blood pressure: A systematic review and critical synthesis of studies on "demand-control-support" and "effort-reward imbalance" models. *Scandinavian Journal of Work, Environment & Health, 40*(2), 109-132.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44*(11), S78-S94.

Heraclides, A., Chandola, T., Witte, D. R., & Brunner, E. J. (2009). Psychosocial stress at work doubles the risk of type 2 diabetes in middle-aged women: Evidence from the Whitehall II study. *Diabetes Care, 32*(12), 2230-2235.

Heraclides, A., Chandola, T., Witte, D. R., & Brunner, E. J. (2012). Work stress, obesity and the risk of Type 2 diabetes: Gender-specific bidirectional effect in the Whitehall II study. *Obesity (Silver Spring), 20*(2), 428-433. doi:10.1038/oby.2011.95

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*(4), 409-426.

Karasek, R., Brisson, C., Kawakami, N., Houtman, I., Bongers, P., & Amick III, B. C. (1998). The job content questionnaire (JCQ): An instrument for internationally comparative assessments of psychosocial job characteristics. *Journal of Occupational Health Psychology, 3*(4), 322-355.

Karasek, R., & Theorell, T. (1990). *Healthy work: Stress productivity and the reconstruction of working life*. New York: Basic Books Inc.

Kivimaki, M., Head, J., Ferrie, J. E., Shipley, M., Brunner, E., Vahtera, J., & Marmot, M. G. (2006). Work stress, weight gain and weight loss: Evidence for bidirectional effects of job strain on body mass index in the Whitehall II study. *International Journal of Obesity, 30*, 982-987.

Landsbergis, P. A., Dobson, M., Koutsouras, G., & Schnall, P. (2013). Job strain and ambulatory blood pressure: A meta-analysis and systematic review. *American Journal of Public Health, 103*(3), e61-e71.

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*(11), S69-S77.

Muthen, L. K., & Muthen, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles: Muthen & Muthen.

Norberg, M., Stenlund, H., Lindah, B., Andersson, C., Eriksson, J. W., & Weinehall, L. (2007). Work stress and low emotional support is associated with increased risk of future type 2 diabetes in women. *Diabetes Research and Clinical Practice, 76*(3), 368-377.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw Hill.

Ohlin, B., Berglund, G., Rosvall, M., & Nilsson, P. M. (2007). Job strain in men, but not in women, predicts a significant rise in blood pressure after 6.5 years of follow-up. *Journal of Hypertension, 25*(3), 525-531.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354-373. doi:10.1037/a0029315

Sanne, B., Torp, S., Mykletun, A., & Dahl, A. A. (2005). The Swedish Demand-Control-Support Questionnaire (DCSQ): Factor structure, item analyses, and internal consistency in a large population. *Scandinavian Journal of Public Health, 33*(3), 166-174.

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). New York: Routledge.

Smith, P. M., Glazier, R. H., Lu, H., & Mustard, C. A. (2012). The psychosocial work environment and incident diabetes in Ontario, Canada. *Occupational Medicine, 62*(6), 413-419.

Smith, P. M., Mustard, C. A., Lu, H., & Glazier, R. H. (2013). Comparing the risk associated with psychosocial work conditions and health behaviours on incident hypertension over a nine-year period in Ontario, Canada. *Canadian Journal of Public Health, 104*(1), e82-e86.

Statistics Canada. (2013). *Canadian Community Health Survey (CCHS) - mental health, user guide microdata files*. Ottawa: Statistics Canada.

Statistics Canada. (2014). *Canadian Community Health Survey (CCHS), mental health public use microdata File, derived variable specifications*. Ottawa: Statistics Canada.

Trudel, X., Brisson, C., & Milot, A. (2010). Job strain and masked hypertension. *Psychosomatic Medicine, 72*(8), 786-793.

van der Doef, M., & Maes, S. (1999). The job demand-control(-support) model and psychological well-being: A review of 20 years of empirical research. *Work & Stress, 13*(2), 87-114.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of Multi-group Confirmatory Factor Analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation, 12*(3), 1-26.

Received: 5.9.2016
Revised: 9.24.2016
Accepted: 9.25.2016