

Factor structure and measurement invariance of a 10-item decisional balance scale:

Longitudinal and subgroup examination within an adult diabetic sample.

Michael A. Pickering, PhD^{1,2}

Ronald C. Plotnikoff, PhD^{1,2,3}

1. Faculty of Physical Education and Recreation, University of Alberta, Canada
2. Alberta Centre for Active Living, University of Alberta, Canada.
3. Centre for Health Promotion Studies, School of Public Health, University of Alberta, Canada.

Corresponding author:

Dr. Michael A. Pickering
Faculty of Physical Education and Recreation
University of Alberta
Edmonton, Alberta
CANADA
T6G 2T4
Telephone: (780) 492-1019
Fax: (780) 492-9579
Email: mpickeri@ualberta.ca

Factor structure and measurement invariance of a 10-item decisional balance scale:

Longitudinal and subgroup examination within an adult diabetic sample

.

ABSTRACT

This study explores longitudinal and subgroup measurement properties of a 10-item, physical activity decisional balance scale, previously published by Plotnikoff, Blanchard, Hotz & Rhodes (2001), within a diabetic sample of Canadian adults. Results indicated that a three-factor measurement model consistently improved model fit compared to the previously published two-factor model. Evidence of configural, metric and scalar measurement invariance, across time and amongst subgroups, suggest the 10-item decisional balance scale is appropriate for investigating associative relationships with other constructs, and for comparing group means of the pros and cons subscales amongst a variety of diabetic population subgroups.

Keywords: physical activity, decisional balance, pros, cons, measurement invariance, confirmatory factor analysis

Decisional balance refers to the perceived “pros” (advantages) and “cons” (disadvantages) of continuing or adopting a new behavior (Plotnikoff et al., 2001; Prochaska & Velicer, 1997). Within theory-based physical activity research, decisional balance is one of four primary components of the Transtheoretical Model (TTM), along with stage of change, self-efficacy and processes of change (Prochaska & Marcus, 1994; Prochaska et al., 1994; Prochaska & Velicer). As such, TTM researchers have frequently included decisional balance as a hypothesized or implied predictor of physical activity behavior (Cox, Stimpson, Poole & Lambur, 2003; Plotnikoff, Brez & Brunet, 2003; Sullum, Clark & King, 2000; Towers, Flett & Seebeck, 2005). Decisional balance has also been included as a psychological outcome or cognitive mediator within TTM-guided, physical activity intervention research (Bock et al., 1997; Plotnikoff, et al., 2007). Furthermore, several investigators whose studies are not explicitly based on TTM, but who focus on improving understanding of physical activity, have included decisional balance as a relevant variable in their research designs (Bopp, Wilcox, Oberrecht, Kammermann, & McElmurray, 2004; Molaison, 2004). In short, decisional balance is a cognitive construct of interest to researchers attempting to better understand psychological aspects of physical activity.

The existence of compelling scientific evidence that physical activity contributes to a wide variety of health benefits (Kravitz, 2007; Warburton, Nicol & Bredin, 2006) has been used as justification for much research aimed at better understanding how psychological factors relate to physical activity (Cox, Stimpson, Poole & Lambur, 2003; Suminski & Petosa, 2006; Towers, Flett & Seebeck, 2005). Diabetics represent one example of a large sub-population (diabetes is estimated to affect 170 million individuals worldwide; Stumvoll, Goldstein & van Haeften, 2005) whose quality of life (Chyun et al., 2003) and physiological mechanisms associated with their disease, can be enhanced through regular physical activity (Herbst, Bachran, Kapellen & Holl, 2006; Herbst et al., 2007; Holt et al., 2007; Kogawa et al., 2007). Given the current high interest in improving our understanding of how cognitions such as decisional balance are associated with and influence physical activity, along with the endemic proportion of the population that is

affected by diabetes, it is not surprising that inquiry aimed at better understanding psychological aspects of physical activity behavior within diabetic populations is beginning to appear in health promotion research literature (Jacobs-van der Bruggen et al., 2007; Plotnikoff, Brez & Brunet, 2003; Plotnikoff, Lippke, Courney, Birkett & Sigal, [in press]).

Despite its central role in TTM and its relevance to contemporary health and physical activity research, the specific measurement properties of instruments designed to assess physical activity-related decisional balance have received only limited empirical attention. One early study by Marcus, Rakowski & Rossi (1992), described item generation procedures and employed a principle component analysis (PCA) with data obtained from a worksite sample, to arrive at a two-factor decisional balance scale that included 10 “pros” items and six “cons” items. The researchers reported Cronbach alpha coefficients of .95 and .79 for pros and cons factors respectively, and indicated that their two-factor solution accounted for 60.4% of variance in the 16 retained items. A more recent study by Plotnikoff et al. (2001), reported measurement properties of a revised version of the Marcus et al. decisional balance scale. The modified version consisted of five pros indicators and five cons indicators. A principle component analysis, using a random sample of 453 working age adults, indicated a two-factor solution accounted for 49.6% of the variance in the 10 items. Cronbach α coefficients were .79 and .71, and two-week test-retest reliability coefficients (r) were .84 and .74, for the pros and cons factors respectively.

Plotnikoff and colleagues (2001) followed their initial scale development efforts with a longitudinal measurement invariance analysis aimed at examining the psychometric robustness of the 10-item decisional balance scale over time (and simultaneous climatic conditions). Their results, using a random sample of 703 Canadian adults between the ages of 18 and 65 who completed the instrument at six-month intervals, suggested the general factor structure and the item-factor loadings did not differ significantly between time points during the 12-month period of the study. Plotnikoff et al. reported that Cronbach α ranged from .77 to .83 for the pros subscale, and from .69 to .72 for the cons subscale.

The reports by Marcus et al. (1992) and Plotnikoff et al. (2001) provide an important starting point for understanding the measurement properties associated with assessing decisional balance related to physical activity in healthy populations. In particular, the longitudinal measurement invariance procedures implemented by Plotnikoff and colleagues provide evidence that the two-factor decisional balance measurement model appears stable in general form, and that the observed item scores scale to the underlying factors in a consistent manner over time. In measurement terminology, these findings suggest “configural” (i.e. equal form) and “metric,” or sometimes labeled “weak,” invariance (i.e. equal factor loadings; Brown, 2006; Meredith, 1993). These two invariance characteristics provide basis for inferring that the general interpretation of the two decisional balance constructs are stable over time (and climatic conditions) for a general population of adult, working age, Canadians. Thus, the configural and metric invariance evidence provided by Plotnikoff et al. supports use of the 10-item decisional balance scale for examining structural relationships (i.e. associations) of perceived pros and cons of physical activity with other constructs of interest (Brown; Steenkamp & Baumgartner, 1998).

Evidence of both configural and metric invariance is fundamental to establishing the overall measurement invariance and conceptual interpretation of an instrument. However, even in combination, both types of invariance do not provide sufficient evidence for making unambiguous comparison of mean levels of the underlying concepts, either over time (Brown, 2006; Chan, 1998), or between population subgroups (Brown; Meredith, 1993; Steenkamp & Baumgartner, 1998). Appropriate comparisons of group means rest on the assumptions of configural and metric invariance, as well as “scalar” or what is sometimes referred to as “strong” invariance (i.e. equal item intercepts; Brown; Meredith). Demonstration of scalar invariance indicates that groups of individuals who have the same value on the latent pros and cons constructs will report similar values on the observed indicators, regardless of subgroup membership (Brown; Rusticus & Hubley, 2006) or assessment time (Brown; Chan). Failure to satisfy the condition of scalar invariance is referred to as an issue of “differential item functioning” (Brown). In regards to the decisional balance scale, existence of differential item functioning implies that for a given level of

pros or cons, members of subgroups (or measures taken at different times) exhibit biased scores on particular items. As a result, the magnitude of observed differences between groups (or times), as exhibited by different mean subscale scores, cannot be unambiguously interpreted as representing actual between-group differences in the underlying pros or cons constructs. Unless scalar invariance exists, observed mean differences might be due to higher or lower scores on biased items, even when the underlying level of perceived pros and cons are similar between the groups (or at different time points).

As pointed out by Horn and McArdle (1992), establishing psychometric properties of an instrument with a varied sample, even one representative of the overall population, does not guarantee identical measurement properties for population subgroups. Therefore, use of the instrument to make unambiguous subgroup comparisons is not justified. Thus, even though the evidence provided by Plotnikoff et al. (2001) supports longitudinal configural and metric invariance for a Canadian adult population of working age, further evidence is necessary to make meaningful decisional balance comparisons involving subgroups of this general population.

Recent research has explored differences in physical activity behavior and in related psychological constructs based on gender (Lochbaum, Bixby & Wang, 2007; Phongsavan, McLean & Bauman, 2007), education (Morrato et al., 2003; Brown, Yore, Ham & Macera, 2005), age (Skelly, Dougherty, Gesler & Soward, 2006; Renner, Spivak, Kwon & Schwarzer, 2007), disease type (Plotnikoff & Lippke et al, 2007), and across time (Nizt & Choy, 2007; Stamatakis, Ekelund & Wareham, 2007). Valid interpretation of research examining decisional balance differences amongst these types of population subgroups, or between time points, will be enhanced by evidence of configural, metric and scalar measurement invariance.

The need for robust evidence that supports unambiguous comparisons of the perceived pros and cons related to physical activity, between population subgroups and from one assessment to another, presents opportunity for further investigation of the measurement properties of the 10-item decisional balance scale published by Plotnikoff et al. (2001). Therefore, our specific objectives in this study are to (1) evaluate the 10-item measurement model previously reported for

the general population, within a sample of Canadian adult diabetics, (2) re-specify the measurement model if empirically and conceptually justified for this population, and (3) assess the configural, metric and scalar invariance properties of the resulting model, longitudinally across three time points, and between several demographic subgroups of our diabetic sample (i.e. sex, age, education and diabetes type).

Method

Participants

A total of 2319 participants (1154 men, 1165 women) took part in this study. The men ranged in age from 18 to 92 years ($M = 61.1$, $SD = 14.0$), and the women ranged in age from 18 to 89 years ($M = 57.8$, $SD = 15.4$). Approximately two-thirds of both the men (69.5%, $n = 797$) and women (64.5%, $n = 743$) reported having been told they had type 2 diabetes. Conversely, just under one-third of both men (28.2%, $n = 323$) and women (32.3%, $n = 372$) reported having type 1 diabetes. Less than 1% of both men ($n = 6$) and women ($n = 7$) indicated they thought they had both, type 1 and type 2 diabetes¹. A few men (1.8%, $n = 21$) and women (2.6%, $n = 30$) indicated they were not sure what type diabetes they had. Less than 1% ($n = 20$) of participants did not respond to the item asking about specific diabetes type. Most participants (76%, $n = 1762$) reported having completed at least a high school education, and 36.5% ($n = 847$) reported having completed college. Less than one quarter (22.8%, $n = 528$) did not have a high school degree, and just over 1% ($n = 29$) did not indicate their highest level of education completed. The majority of respondents reported being of Canadian ethnic origin (76.5%, $n = 1775$), and 14.2% ($n = 329$) reported being of European origin. The remaining 8.4% ($n = 195$) of participants who responded to the ethnicity item selected one of the following: Arab (4), Asian (51), African (10), Aboriginal (18), Latin-S. American (4) or Other (108). Less than 1% ($n = 20$) of participants did not reply to the ethnicity item, or reported they did not know their ethnic origin. Only 8.2% of study participants indicated they had never been married, over three quarters (76.8%, $n = 1728$) reported either being married or in a common-law partnership, and 16.6% ($n = 385$) indicated they were separated, divorced or widowed. Less than 1% ($n = 16$) did not respond to the marital status item.

Over one third (36.2%, $n = 841$) of study participants reported an annual family income of less than \$40,000, another 36% ($n = 835$) reported annual income between \$40,000, and \$79,999, and 17% ($n = 394$) reported annual income of \$80,000 or more. The percentage of respondents who did not to respond to the family income question (10.7%, $n = 249$) was considerably higher than was non-response for other items.

Measures

The measures used in these analyses consisted of demographic variables, including those described above, and the 10-item decisional balance scale previously reported by Plotnikoff et al. (2001). Study participants responded to both pros and cons items using a 5-point Likert-type scale consisting of the following possible choices: (1) not at all, (2) a little, (3) somewhat, (4) quite a lot, and (5) very much. Appendix A lists the specific wording of the 10 decisional balance items. A complete description of the procedures used to develop the 10-item decisional balance scale has been reported by Plotnikoff and colleagues (2001) and Plotnikoff (2002), and has been commented on by Cardinal (2002).

In addition to the demographic items, and the ten items making up the decisional balance scale, the survey also included items assessing current physical activity level (Godin, 1997), personality dimensions (Goldberg, 1992) and several cognitive constructs related to behavior change theories (e.g. the Trans-theoretical Model, the Theory of Planned Behavior, and Protection Motivation Theory). However, these additional data were not used in these analyses.

Procedure

The data for these analyses come from the Alberta Longitudinal Exercise and Diabetes Research Advancement (ALEXANDRA) study. The ALEXANDRA study examined social cognitive determinants of physical activity of individuals with type 1 and type 2 diabetes over an 18 month period (i.e. baseline, six, and 18 months). Detailed sampling procedures and response rates for ALEXANDRA data have been reported by Plotnikoff et al. (2006). Briefly, the sample consisted of individuals recruited; (1) from the Canadian Diabetes Association registry, (2) via a random digit-dialing method (RDD) developed by a university-based population research lab, and

(3) through identification of diabetic family members of those contacted via RDD, not living in the RDD contacted household. To avoid duplicate responses, participants were asked to complete the baseline survey only once. Postage-paid return questionnaires and consent forms accompanied a regular mailing of the Canadian Diabetes Association newsletter. The same materials were mailed to individuals contacted via RDD and to the selected diabetic family members, after initial telephone contact. A university ethics review board approved all study procedures.

Model evaluation

There were two main phases involved in our investigation of the psychometric properties of the 10-item decisional balance scale. In the first phase (i.e., the “calibration” phase) we submitted a portion of randomly selected cases ($n = 743$ cases, 33.3% of the study sample) to a confirmatory factor analysis of the two-factor measurement model previously reported by Plotnikoff et al. (2001). Our initial aim was to evaluate the fit of the previously published model to our diabetic sample (i.e. study objective 1). Upon consideration of overall model fit and individual item-factor loadings, we re-specified a revised measurement model on the basis of empirical results from covariance model testing² and exploratory factor analysis, while remaining mindful of conceptual interpretation (i.e. study objective 2). The second phase, (i.e. the “validation” phase) consisted of (1) a confirmatory evaluation of the re-specified model that resulted from the calibration phase, (2) a longitudinal invariance analysis, and (3) a series of subgroup invariance analyses (i.e. study objective 3). All analyses in the validation phase were performed using the remaining cases not selected for the calibration procedures.

All model evaluations, including both initial confirmatory analysis of the two-factor model reported by Plotnikoff, et al. (2001), exploratory covariance modeling of the re-specified model, and all confirmatory, longitudinal and subgroup invariance testing of the new model, were performed using the AMOS version 16.0 software, distributed by SPSS, Inc., installed on a personal computer running the Microsoft Windows XP operating system. Descriptive analyses and exploratory factor analyses were performed with SPSS version 16.0, running on the same platform.

For all invariance analyses, model adequacy was first examined separately for the subgroups. If the model appeared to fit the data reasonable well for the individual sub samples, assessment of invariance between the groups continued by sequentially testing a series of progressively restrictive and nested models in the following order: (1) Model a - equal form, (2) Model b - equal factor loadings, (3) Model c – equal factor variances, (4) Model d - equal factor co-variances, (5) Model e - equal indicator intercepts, and (6) Model F - equal latent means. Our rationale for the model testing sequence was that, Models a, b and e specifically assess configural, metric and scalar measurement invariance respectively. Models c, d and f compare subgroup distributional properties of the constructs (I.e. Do the sub samples exhibit similar amounts of (a) pros and cons construct variability, (b) pros and cons construct co-variability, and (c) average levels of perceived pro's and con's associated with physical activity?). Appropriate comparison of both variability and co-variability assume both configural and metric invariance between subgroups (i.e. Models a and b). Unambiguous comparison of subgroup means (i.e. Model f) is also dependent upon configural and metric invariance. Plus, it assumes consistent item functioning between groups (i.e. scalar invariance), which we explored by comparing item intercepts via Model e (Brown, 2006; Cheung & Rensvold, 2002).

Specification of the basic measurement models involved setting one item loading for each factor to 1.0, to scale the latent factor metric. Remaining item loadings on each primary factor were freely estimated, and were restricted to equal zero for other factors. Covariances between the factors were freely estimated. Covariances between all item uniqueness values (i.e. “error”) were set to zero. Subsequent models in the invariance procedure were specified by systematically adding parameter constraints according to recommendations made by Brown (2006).

Three goodness-of-fit indices were used to evaluate overall model fit at every stage of the analyses. For the first two incremental fit indices, the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI), values above .90 and .95 were judged as indicating adequate and good fit respectively (Hu & Bentler, 1999; Vandenberg & Lance, 2000). The root mean square error of approximation (RMSEA), an indicator of absolute fit, was the third fit index. Values of RMSEA

less than .08 were judged to indicate reasonable overall model fit (Browne & Cudeck, 1993), and values below .05 were considered evidence of good model fit to the data (Hu & Bentler, 1999). The Akaike Information Criteria (AIC; Anderson, Burnham & Thompson, 2000), a predictive fit index, was computed for both models in the calibration phase of the study.

Invariance between hierarchically nested models is commonly assessed with chi-square difference tests (χ^2_{diff} ; Cheung & Rensvold, 2002; Rusticus & Hubley, 2006). However, methodologists have acknowledged that χ^2_{diff} is influenced by sample size (Brannick, 1995; Cheung & Rensvold, 2002; Kelloway, 1995). Based on a simulation analysis of 20 goodness-of-fit indices, Cheung & Rensvold (2002) proposed using changes in CFI (CFI_{diff}) of greater than .01 as an alternate criterion for evaluating multi-group measurement invariance. Conversely, French & Finch (2006) concluded from their own simulation study that chi-square difference tests often yield reasonable outcomes regarding group invariance testing of overall model fit, and offer satisfactory protection against type-1 error. Our strategy in these analyses was to report both, χ^2_{diff} and CFI_{diff} values, and to note relevant similarities and differences where appropriate. In consideration of sample size sensitivity, our χ^2_{diff} tests were guided by a p-value cut-off of .01.

Sub samples

calibration vs. validation. Prior to performing any analyses the overall sample was randomly divided into a calibration sample (n = 773) and a validation sample (n = 1546). The calibration sample was used to evaluate how well the previously reported two-factor model fit a diabetic sample, and for the exploratory re-specifying of a better fitting model. The validation sample was used for confirmatory evaluation of the revised model, the longitudinal invariance analysis, and was further subdivided into sub samples for the subgroup invariance analyses. All calibration analyses and subgroup analyses were performed using baseline data only.

cross-validation of revised measurement Model. Before examining longitudinal invariance, and subgroup invariance properties based on demographics and diabetes type, the validation sample was randomly divided into two sub samples (n = 743 each) and submitted to the full measurement invariance procedure. Because differences between these two validation sub

samples, and between these sub samples and the calibration data, reflected only random fluctuations (i.e. all three sub samples were randomly selected from the overall study sample), we expected the revised model to fit the validation sub samples well, and for these data to fulfill all requirements of full measurement invariance. Likewise, we expected distributional properties (i.e. means, variances, and covariances) of the sub samples to be similar to each other, and to the calibration data. This initial cross-validation served as a robust confirmatory analysis of the revised measurement model specified in the calibration phase, based on two presumably similar, but independent validation samples.

gender. In order to examine measurement invariance between males and females, the validation sample was divided into two sub samples based on sex. The resulting groups consisted of 768 females with a mean age of 57.4 years ($SD = 15.6$) and 778 males with a mean age of 61.1 years ($SD = 14.1$). Of the females, 31% indicated they had type 1 diabetes and 65% indicated they had type 2 diabetes. Of the males, 28% reported they had type I, and 69% reported type 2.

diabetes type. In order to examine measurement invariance between type 1 and type 2 diabetics, the validation sample was divided into two sub samples based on diabetes type. The resulting groups consisted of 456 type 1 individuals with a mean age of 50.7 years ($SD = 17.2$), and 1037 type 2 individuals with a mean age of 62.8 years ($SD = 12.0$). For type 1, 52% were females and 48% were males. For type 2 individuals, 48% were females and 52% were males.

education. For purposes of examining measurement invariance between level of education, participants in the validation sample were divided into two categories: (1) 685 individuals with a high school education or less. This group had a mean age of 63.1 years ($SD = 13.0$) and consisted of 51.1% men and 48.9% women; and (2) 842 individuals with more than a high school education. This group had a mean age of 56.1 years ($SD = 15.8$) and consisted of 49.9% men and 50.1% women.

age groups. For purposes of examining measurement invariance across age, participants in the validation sample were divided into three age categories as follows: 18 to 50 years (170 men,

243 women), older than 50 but younger than 70 years (346 men, 329 women), and 70 years and older (262 men, 196 women).

Missing Data

For the calibration sample we used the default, maximum likelihood estimated missing data method implemented in the AMOS software (Arbuckle, 2006). One advantage of this approach, over alternatives such as list-wise deletion, pair-wise deletion or mean imputation, is that it assumes only the data are missing at random, while the alternatives just mentioned rely on the more strict assumption of data being missing completely at random (Arbuckle).

For the validation samples used in the invariance analyses we initially desired the option of examining modification indices in the case of poorly fitting models³. However, the AMOS modification index computation algorithms require complete data (Arbuckle, 2006). In order to address this requirement, we took advantage of the AMOS software's Bayesian imputation capability to produce estimated values for missing data cells. Because Bayesian imputation estimates vary from one completed dataset to another, we created 10 different complete datasets of the baseline, six-month, and 12-month data. We then computed a final dataset for each assessment, with the missing data cells replaced by the mean of the 10 imputed cell estimates. It should be noted that this is not identical to the process typically used in "multiple imputation," whereby imputed datasets are analyzed separately, and the resulting estimates are then combined into a single set of results (Arbuckle). To do so for this study would have required analyzing a total of over 500 separate models. Our revised procedure involved performing 50 analyses on the complete dataset, plus examination of follow-up models for purpose of identifying specific sources of invariance.

Results

Preliminary analysis

Due to the large number of sub-samples and models involved in these analyses, space limitations do not permit presentation of subgroup- and assessment time-specific inter-item correlation matrices, item statistics, nor individual parameter estimates for each model. (A

complete set of tables with this information is available from the corresponding author.)

Preliminary examination of item distributions indicated that pros items exhibited negative skewness, with indices ranging from 0.66 to -1.08 , and cons items exhibited positive skewness, with indices ranged from 0.71 to 1.76. Because, these data do not fully meet normality assumptions we specified unbiased covariance matrices as input to the AMOS maximum likelihood estimation procedure, in order improve the robustness of parameter estimates and associated standard errors (Brown, 2006).

Two-factor CFA and subsequent model re-specification (calibration)

With respect to our first study objective, Figure 1 displays confirmatory factor analysis results for the original two-factor measurement model of the 10-item decisional balance scale (Plotnikoff et al., 2001), with standardized item-factor loadings, explained percent indicator variance, and construct correlation, along with the select goodness of fit indices. Figure 2 shows the same information for a revised three-factor model.

Brown (2006) explains how the three-factor model is nested under the two-factor model, but with two fewer degrees of freedom. Therefore, sample size issues notwithstanding, the difference in overall model fit can be evaluated using the traditional chi-square difference test. For these calibration data, $\chi^2_{\text{diff}}(2) = 84.168$, $p < .001$, indicating that the three-factor model fit the data significantly better than the two-factor model. Likewise, $\text{CFI}_{\text{diff}} = .033$ and $\text{TLI}_{\text{diff}} = .050$, both suggesting substantial improvement in goodness-of-fit in favor of the revised model. The AIC and the RMSEA are both considerably lower for the three-factor model than for the two-factor model, again confirming better overall fit to the data by the three-factor model.

All estimated item-factor loadings for both models were highly significant (i.e. $p < .001$). However, the percentages of variance accounted for in items “con04t1” and “con05t1” were over five times higher for the three-factor model than for the two-factor model (i.e. 31% vs. 6%, and 43% vs. 8%, for fourth and fifth cons items respectively).

Although our re-specification was guided primarily by the low observed loadings of the of the fourth and fifth cons items in the original model, we also submitted the baseline calibration

data to two exploratory factor analyses (EFA) using maximum likelihood factor extraction and oblique rotation (SPSS v16 oblimin). Unlike the covariance models, the EFA models did not pre-specify loading patterns, and freely estimated cross-loadings of items on all factors. The purpose for subjecting the calibration data to an EFA was to compare the structure of SPSS generated EFA solutions (based on ML extraction and Oblimin rotation) to the measurement models being considered as a result of the prior covariance analyses (i.e the original two-factor model, and a potential three-factor model that distinguished the fourth and fifth cons items). To do so, we first restricted the EFA to extract two factors (i.e., the same number of factors in the original measurement model), which resulted in a solution identical in form (disregarding EFA cross-loadings) to the model previously published by Plotnikoff et al. (2001). Next, we specified the EFA to extract factors with eigenvalues greater than 1.0, which resulted in a three factor solution resembling (again, disregarding EFA cross-loadings) our intended re-specified covariance model that would distinguish the first three cons items from the fourth and fifth. Specific factor loadings for both EFA models are displayed in Table 1. The respective chi-square fit indices for the two- and three-factor EFA solutions were $\chi^2(26) = 223.2, p < .001$ and $\chi^2(18) = 24.9, p = .127$, suggesting fit of the three-factor EFA solution was preferable to the EFA solution restricted to extracting only two factors. Likewise, the total variance accounted for in the set of 10 pros and cons items was greater for the three-factor solution (69.33%) than it was for the two-factor solution (59.50%).

The empirical evidence favoring the revised three-factor measurement model was corroborated by our subjective interpretation of the item content. Specifically, the first three cons items appeared to tap perceived disadvantages related to time/responsibility conflicts, while the wording of the fourth and fifth cons items related to perceived risk of injury or unfavorable social judgment. Thus, the empirical results combined with our subjective interpretation, led us to choose the three-factor model as the basis of the subsequent longitudinal and subgroup invariance analyses. Based on the three-factor model, the calibration data exhibited coefficient α 's of 0.85 and 0.79 for the pros factor and the time/responsibility cons factor respectively, and the two items

contributing to the remaining cons factor exhibited a modest positive correlation, $r = 0.36$. For the EFA solution, the item variance explained by each of the three factors was 34.0% for the pros factor, 24.5% for the time/responsibility cons factor, and 10.8% for the second cons factor.

Cross-validation of revised three-factor model with two independent samples

The top panel of Table 2 displays goodness of fit indices for the initial cross-validation analyses of the revised three-factor measurement model. Both incremental fit indices and the RMSEA, indicated the revised measurement model fit both independent validation samples well. Likewise, when the model was fit to both validation sub samples simultaneously (i.e., Model a – equal form), the fit was also very good. Model a served as the comparison model for subsequent invariance testing.

As shown in Table 2, both CFI_{diff} values for Models b and e were well below the invariance cut-off value of .01, suggesting constraints associated with metric and scalar invariance did not significantly degrade the overall fit of the model to the data. Likewise, the χ^2_{diff} difference values for Models b and e were also non-significant ($p > .01$).

The results of the measurement invariance models (i.e. Models a, b & e) justified comparison of distributional properties (i.e. variances, co-variances and latent means) of the sub-samples. For the three models making these comparisons (i.e. Models c, d & f respectively), the CFI_{diff} values were well below the .01 criteria, and all three χ^2_{diff} values were non-significant ($p > .01$). Thus, the pros and cons constructs showed similar amounts of variability, co-variability, and mean levels in the cross-validation samples. As expected, the three-factor model satisfied all conditions of measurement invariance, and invariance of distributional properties.

Longitudinal invariance models

From Table 2, CFI_{diff} values from both Model b and Model e suggest longitudinal measurement invariance. However, for Model c (i.e. equal factor variances) $CFI_{diff} = .020$, suggesting for at least one of the pros or cons constructs, variability is not consistent across time. Constraining the variance for each construct alone also resulted in substantially poorer fitting models than Model a (i.e. $\chi^2_{diff-pros} (16) = 200.39, p < .001$; $\chi^2_{diff-cons1} (16) = 184.48, p < .001$;

$\chi^2_{\text{diff-cons2}}(16) = 76.33, p < .001$). Examination of the estimated variance parameters in Model b (i.e. equal loadings only, with no constraints on variance) indicated a consistent and substantial pattern of reduced variability for all three constructs with each subsequent assessment.

In contrast to the CFI_{diff} -based results, the χ^2_{diff} tests suggested that Model b exceeded the invariance criterion ($\chi^2_{\text{diff}}(14) = 29.32, p < .01$). Follow-up analyses revealed that releasing the equal loading constraint on the second pros item significantly improved the overall model fit, as compared to Model b, where all item-factor loadings were constrained to be equal across assessments ($\chi^2_{\text{diff}}(2) = 16.71, p < .001$). Further investigation revealed that only the loading at the third assessment appeared substantially different for this item. Specifically, when item-factor loadings for the second pros item were constrained to be the same at the first two assessments, but freely estimated for the third, the overall model fit no longer was significantly different from Model a ($\chi^2_{\text{diff}}(12) = 12.75, p > .20$), but was significantly better than Model b ($\chi^2_{\text{diff}}(1) = 16.33, p < .001$). The item-factor loading for pro02t1 at the first two assessments was 0.11 units greater in magnitude than it was at the third assessment.

Technically, metric non-invariance precludes unambiguous investigation of further measurement and distributional property invariance. However, recognizing the relatively small magnitude of the factor loading scale difference exhibited for only one item, and for only one time point, it is worth noting that χ^2_{diff} -based evaluation of the fit of Model c, also suggests construct variance differences amongst the time points ($\chi^2_{\text{diff}}(20) = 380.07, p < .001$).

Subgroup invariance models

gender invariance. From Table 3, it is evident that all CFI_{diff} values, and associated χ^2_{diff} values, for Models b and e were below criteria that would indicate substantial measurement differences between males and females for either factor loadings or indicator intercepts, suggesting both configural and metric invariance across gender. Likewise, fit of Models c and d were not significantly different from the baseline equal form model, providing evidence of similar variability and co-variability. However, CFI_{diff} for Model f (i.e. equal latent means) exceeded the .01 criteria recommended by Cheung & Rensvold (2002). Likewise, for Model f, $\chi^2_{\text{diff}}(17) =$

88.76, $p < .001$. Thus, constraining the latent pros and cons mean values for males and females to be equal, significantly degraded the fit of the model. Inspection of parameter estimates of Model e, where the latent means were not constrained to equality, indicated that females reported slightly higher mean levels than males on all three latent constructs. The magnitudes of these differences were 0.17, 0.12 and 0.24 for the pros, cons1 and cons2 constructs respectively. Follow-up tests indicated that constraining equal means for males and females on any one of the constructs alone significantly degraded the overall fit of the model as compared to Model a, (i.e. $\chi^2_{\text{diff-pros}} (15) = 41.73$, $p < .001$; $\chi^2_{\text{diff-cons1}} (15) = 35.43$, $p < .01$; $\chi^2_{\text{diff-cons2}} (15) = 71.83$, $p < .001$).

diabetes type invariance. None of the CFI_{diff} values in Table 3 suggest either measurement differences or distributional differences in the pros and cons constructs between type 1 and type 2 diabetics. However, for Model e, $\chi^2_{\text{diff}} (14) = 37.56$, $p < .001$, suggesting the possibility of some indicator bias. Examination of the estimated item intercepts for both groups indicate that the largest intercept difference between type 1 and type 2 diabetics occurred for item “pros05t1.” The magnitude of this difference was 0.30, with type 2 diabetics reporting the higher intercept value. When the intercepts on this item were freely estimated, but all other intercepts were constrained to equality across diabetes types, the overall fit of the model was not significantly degraded as compared to Model a ($\chi^2_{\text{diff}} (13) = 13.18$, $p > .30$), and the overall model fit significantly improved as compared to Model e (i.e. full restriction of all item intercepts; $\chi^2_{\text{diff}} (1) = 24.38$, $p < .001$). Thus, based on χ^2_{diff} – based evaluation, the fifth pros item appears to exhibit slight, but statistically significant, positive bias for type 2 diabetics.

education level invariance. None of the CFI_{diff} values in Table 3 suggest either measurement differences, or distributional differences in the pros and cons constructs between individuals with more than a high school education, and those with a high school education or less. However, for Model e, $\chi^2_{\text{diff}} (14) = 36.07$, $p < .01$, suggesting potential for some indicator bias. Follow-up comparisons that relaxed the equality constraint on the pro05t1 item significantly improved the overall model fit as compared to Model e ($\chi^2_{\text{diff}} (1) = 10.91$, $p < .01$). Likewise, fit of this revised model was not significantly degraded compared to Model a, the equal form model

($\chi^2_{\text{diff}}(13) = 26.26, p > .01$). The magnitude of the difference in intercepts between the groups was 0.17, with the higher educated group exhibiting the lower intercept for the pro05t1 item.

age invariance. From Table 3, CFI_{diff} values from both Model b and Model e suggests measurement invariance across the three different age groups. CFI_{diff} values from Models c and d also indicate that all three age groups exhibited similar amounts of construct variability and covariability. Only Model F exhibited a CFI_{diff} value greater than .01, indicating that constraining the means of the constructs to be equal for all three age groups substantially degraded the fit of the model compared Model a. The top panel in Table 4 displays estimated latent mean differences between the older two age groups, and the younger participants, in a model without latent mean equality constraints. The second panel displays mean differences from a similar model where only the means for the pros and cons2 constructs are constrained to be equal for the younger two age groups. The fit of this model is not significantly different than Model a ($\chi^2_{\text{diff}}(30) = 55.80, p > .01$). However, further equality constraints do result in significant degradation of model fit.

From Table 3., although the CFI_{diff} values suggest a difference only in latent means amongst age groups, χ^2_{diff} value-based criteria suggest there may be some difference in covariability of the pros and cons constructs amongst age groups (i.e. Model d), and perhaps some indicator bias as well (i.e. Model e). Further follow-up analyses indicated that restricting the covariance between the pros construct and the cons2 construct did not significantly degrade model fit compared to Model a ($\chi^2_{\text{diff}}(22) = 35.96, p > .01$), nor did equality constraints between the older two age groups, on the covariance between pros and cons1 and the two cons constructs ($\chi^2_{\text{diff}}(24) = 36.30, p > .01$). However, adding equality constraints amongst all three age groups to either the covariance between pros and cons1 ($\chi^2_{\text{diff}}(24) = 45.81, p < .01$), or the covariance between the two cons constructs ($\chi^2_{\text{diff}}(24) = 44.63, p < .01$) both resulted in significantly worse model fit than exhibited by Model a. Correlations shown in the bottom panel of Table 4 suggest the slight negative relationship between perceived pros and cons1 appears to lessen with increasing age, while the small positive correlation between the two cons constructs was slightly stronger for the older two age groups, than it was for the participants 50 years of age and younger.

Regarding potential indicator bias, no individual intercept constraints resulted in χ^2_{diff} values indicating significantly poorer fit than the baseline Model a. When intercept equality constraints were placed on the pros03t1 and pros05t1 items simultaneously $\chi^2_{\text{diff}}(18) = 36.03$, $p < .01$, suggesting poorer fit than Model a, and therefore, the possibility of slight indicator bias between age groups for these items. Examination of unconstrained intercept values for pros03t1, indicated that the intercept of the youngest age group was .19 units and .16 units less than for the middle and oldest age groups respectively. For the pros05t1 item, the intercept of the middle age group was .16 units and .18 units more than for the youngest and oldest age groups respectively. Thus, these data suggest slight negative bias on pros03t1 for respondents 50 years of age and younger, and slight positive bias on pros05t1 for respondents between 50 and 70 years of age.

Discussion

Our findings offer several insights into the measurement and distributional properties of the 10-item decisional balance scale when used with Canadian adult diabetics. With regards to the most appropriate form of the measurement model (i.e., our first and second study objectives), we observed substantial improvement in overall model fit by revising the previously published two-factor measurement model (Plotnikoff et al., 2001) in a manner that distinguished the first three cons items from the remaining two, resulting in a three-factor model. All empirical indices of model fit were substantially improved, as were the individual loadings for the fourth and fifth cons items. Furthermore, confirmatory cross-validation of the re-specified three-factor model, with two independent samples, satisfied all requirements of measurement and distributional property invariance, suggesting that the calibration process used to arrive at the three-factor solution resulted in a structurally valid model for this population. Although a three-factor decisional balance model has not been examined in previous literature, it is worth noting that both the overall model fit, and pattern of item loadings, observed in our initial two-factor CFA were similar to those reported by Plotnikoff et al. Specifically, the fourth and fifth cons items exhibited substantially lower relationship with the cons construct than did the first three items.

Substantively, the wording of the first three cons items appears to reflect perceived conflicts of time and/or social responsibilities. The relationship between the two latent cons constructs was positive, but moderate (i.e. $r = 0.41$. from Figure 2), and neither was related substantially with the pros construct. Based on the 10 items in this instrument, it appears that adult diabetics distinguish physical activity-related cons related to time and social responsibility conflicts from those associated with appearing awkward or enduring excessive financial costs. Again, this differentiation has not been conceptualized in prior decisional balance research. However, several researchers exploring perceived barriers to physical activity have reported perceived “lack of time” (Dutton et al., 2005), and “lack of time due to family and employment obligations” (Mancuso et al., 2006) as frequently identified perceived barriers. Interestingly, one recent study that purports to validate a new scale assessing barriers to physical activity in diabetics, suggests a single factor model, but does not include any items that specifically address perceived time or social responsibility conflict (Dube et al., 2006). Unfortunately, even though time and social responsibility are barriers to physical activity commonly cited by adults, this new instrument does not address whether such concerns are part of the common dimension of barrier perceptions exhibited by the items that are included in the scale. Our findings alone, do not inform us to what extent this type of conceptual differentiation is specific to our sample of diabetics, or might be generalized to a wider population of adults. Future research is needed with respect to this decisional balance conceptual issue.

It is also important to note that distinguishing perceived cons related to time and responsibility conflicts from other types of cons, resulted in constructs with only three and two indicators respectively. If this conceptual distinction is maintained in future substantive research, it will be beneficial to identify additional items for use in assessing each dimension.

With regards to our third study objective (i.e. assessing measurement and distributional invariance properties), by and large, the three-factor measurement model exhibited strong psychometric properties, both across time and between population sub-groups. Using the CFI_{diff}

criteria suggested by Cheung & Rensvold, (2002), the three-factor model satisfied all conditions of configural, metric, and scalar invariance, for both longitudinal and all sub-group evaluations.

Regarding distributional properties, females exhibited slightly higher mean levels of both perceived pros and cons than did males, but magnitudes of these differences were relatively small (i.e. ranging from 0.12 to 0.24 units on a five-point rating scale). Likewise, the oldest study participants reported slightly lower levels of both pros and cons than did the other two age groups, and the middle age group reported slightly lower levels of cons related to time and responsibility conflicts than did the youngest age group. Thus, it seems adult diabetics may perceive both fewer benefits, and fewer disadvantages, to adopting or continuing physical activity, as they get older. Finally, variability of all three constructs decreased with each subsequent assessment.

Results of the evaluations based on traditional χ^2_{diff} tests suggest possibility of slight measurement invariance. These analyses suggested the model did not meet the strictest requirements (i.e. non significant χ^2_{diff} values) of scalar invariance for some longitudinal and subgroup comparisons. Specifically, there appeared to be a slight amount of indicator bias between diabetes types, education level, and age groups. These differences are important to be aware of, but should be considered with respect to their relative magnitude, and within context of the acknowledged sensitivity of χ^2_{diff} to sample size. For example, the strongest potential indicator bias occurred for the pros05t1 item, when comparing type 1 diabetics to type 2. The intercept for this indicator differed between subgroups, by .30 units on the five-point response scale. Given that the pros subscale score is typically computed by averaging responses to all five indicators, a reliable bias of .30 units for a single indicator will result in a bias of only .06 units for the pros subscale score. Thus, unless a bias of such magnitude is substantively meaningful, there appears minimal risk associated with validly interpreting findings that compare mean levels of perceived pros associated with physical activity, between Canadian adults with type 1 vs. type 2 diabetes. Likewise, the bias exhibited for the fifth pros item also presents minimal concern for comparisons between those with more than a high school education to those with less.

One methodological limitation to this study is that our subgroup invariance analyses are based on re-dividing the validation sample for each analysis. Thus, the results of each analysis are not completely independent from each other. Furthermore, we have considered only “main effects” types of invariance. We have not attempted to assess measurement differences based on the numerous possible combinations of subgroup membership.

Despite these limitations, our findings suggest the 10-item decisional balance scale offers a valid tool for assessing perceived pros and cons associated with physical activity, by Canadian adult diabetics. Further research is needed exploring population-related external validity issues, and additional scale development work identifying additional items will likely enhance properties of the cons subscales. Overall however, these findings support the use of the 10-item decisional balance scale as a valid tool for assessing pros and cons associated with physical activity, at least for populations and several sub populations of adult Canadian diabetics.

References

- Anderson, D.R., Burnham, K.P. & Thompson, W.L. (2000). Null hypothesis testing: Problems, prevalence, and an alternative. *Journal of Wildlife Management*, 64, 912-923.
- Arbuckle, J.L. (2006) Amos 7.0 user's guide. Spring House, PA: Amos Development Corporation.
- Bock, B.C., Albrecht, A.E., Traficante, R.M., Clark, M.M., Pinto, B.M., Tilkemeier, P. and Marcus, B.H. (1997). Predictors of exercise adherence following participation in a cardiac rehabilitation program. *International Journal of Behavioral Medicine*, 4(1), 60-75.
- Bopp, M., Wilcox, S., Oberrecht, L., Kammermann, S. & McElmurray, C.T. (2004). Correlates of strength training in older rural African American and Caucasian women. *Women and Health*, 40(1), 1-20.
- Brannick, M.T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, 16, 201-213.
- Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Brown, D.R., Yore, M.M., Ham, S.A. & Macera, C.A. (2005). Physical activity among adults ≥ 50 yr with and without disabilities, BRFSS 2001. *Medicine and Science in Sports & Exercise*, 37(4), 620-629.
- Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage.
- Cardinal, B.J. (2002). Decisional balance scale for exercise behavior: A brief commentary with guidelines for item construction. *Measurement in Physical Education and Exercise Science*, 6(4), 249-253.
- Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS)

- and multiple indicator latent growth modeling (MLGLM). *Organizational Research Methods*, 1(4), 421-483.
- Cheung G.W. & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Chyun, D., Khuwatsamrit, K., Katten, D., Price, W., Davey, J., Inzucchi, S. & Melkus, G. (2003). Self-management of type 2 diabetes (T2DM) in adults screened for coronary artery disease (CAD): impact on quality of life (QOL). *Diabetes*, 52(6), A410-A411.
- Cox, R.H., Stimpson, T.S., Poole, K.P. & Lambur, M.T. (2003). Physical activity characteristics of potential participants in nutrition and fitness programs based on stages of change, self-efficacy, and decisional balance. *Family and Consumers Sciences Research Journal*, 31(4), 361-377.
- Dube, M.C., Valois, P, Prud'homme, D. Weisnagel, S.J. & Lavoie, C. (2005). Physical activity barriers in diabetes: Development and validation of a new scale. *Diabetes research and clinical practice*, 72, 20-27.
- Dutton, G.R., Johnson, J., Whitehead, D., Bodenlos, J. & Brantley, P.J. (2005). Barriers to physical activity among predominantly low-income African American patients with type 2 diabetes. *Diabetes Care*, 28(5), 1209-1210.
- French, B.F. & Finch, W.H. (2006). Confirmatory factor analytic procedures for determination of measurement invariance. *Structural Equation Modeling*, 13(3), 378-402.
- Godin, G. & Sheppard, R. (1997). A simple method to assess exercise behavior in the community. *Canadian Journal of Applied Sport Science*. 10, 141-146.
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4, 26-42.
- Herbst, A., Kapellen & Holl, 2006

- Herbst, A., Kordonouri, O., Schwab, K., Schmidt, F. & Holl, R.W. (2007). Impact of physical activity on cardiovascular risk factors in children with type 1 diabetes. *Diabetes Care*, 30(8), 2098-2100.
- Holt, H.B., Wild, S.H., Wareham, N., Ekelund, U., Umpleby, M., Shojaee-Moradie, F., Hold. R.I.G., Phillips, D.I. & Byrne, C.D. (2007). Differential effects of fatness, fitness and physical activity energy expenditure on whole-body, liver and fat insulin sensitivity. *Diabetologia*, 50, 1698-1706.
- Horn, J.L. & McArdle, J.J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Journal of Experimental Aging Research*, 18, 117-144.
- Hu, L.T. & Bentler, P.M. (1999). Cut-off criteria for fit indices in covariance structure analysis: Conventional criteria vs. new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jacobs-van der Bruggen., M.A., Bos, G., Bemulmans, W.J., Hoogenveen, R.T., Vijgen, S.M. & Baan, C.A. (2007). Lifestyle interventions are cost effective in people with different levels of diabetes risk. *Diabetes Care*, 30(1), 128-134.
- Kelloway, K.E. (1995). Structural equation modeling in perspective. *Journal of Organizational Behavior*, 16, 215-224.
- Kogawa Sato, K., Hayashi, T. Kambe, H., Nakamura, Y. Harita, N., Endo, G. & Yoneda, T. (2007). Walking to work is an independent predictor of incidence of type 2 diabetes I Japanese men. *Diabetes Care*, 30(9), 2296-2298.
- Kravitz, L. (2007). The 25 most significant health benefits of physical activity & exercise. *IDEA Fitness Journal*, October, 54-63.
- Lochbaum, M.R., Bixby, W.R. & Wang, C.K. (2007). Achievement goal profiles for self-report physical activity participation: Differences in personality. *Journal of Sport Behavior*, 30(4), 471 – 490.

- Mancuso, C.A., Sayles, B.A., Robbins, L., Phillips, E.G., Ravenell, K., Duffy, C., Wenderoth, S. & Charlson, M.E. (2006). Barriers and facilitators to healthy physical activity in asthma patients. *Journal of Asthma*, 43, 137-143.
- Marcus, B.H., Rakowski, W. & Rossi, J.S. (1992). Assessing motivational readiness and decision making for exercise. *Health Psychology*, 11, 257-261.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525-543.
- Molaison, E.F. (2004). Linking decisional balance to participation in physical activity in college students. *Journal of The American Dietetic Association*, August, A57.
- Morrato, E.H., Hill, J.O., Wyatt, H.R., Ghushchyan, V. & Sullivan, P.W. (2003). Physical activity in U.S. adults with diabetes and at risk for developing diabetes, 2003. *Diabetes Care*, 30(2), 203-209.
- Nitz, J.C. & Choy, N.L. (2007). Changes in activity level in women aged 40-80 years. *Climacteric*, 10, 408-415.
- Phongsavan, P., McLean, G. & Bauman, A. (2007). Gender differences in influences of perceived environmental correlates on recommended level of physical activity among New Zealanders. *Psychology of Sport and Exercise*, 8, 939-950.
- Plotnikoff, R.C. (2002). The development of social-cognitive measures in the exercise domain: Issues and challenges. *Measurement in Physical Education and Exercise Science*. 6(4), 255-261.
- Plotnikoff, R.C., Blanchard, C., Hotz, S.B. & Rhodes, R. (2001). Validation of the decisional balance scales in the exercise domain from the Transtheoretical Model: A longitudinal test. *Measurement in Physical Education and Exercise Science*, 5(4), 191-206.

- Plotnikoff, R.C., Brez, S. & Brunet, S. (2003). Are exercise social-cognitive factors and behaviors different for adults with diabetes? A randomized community sample. *Psychology, Health & Medicine*, 8(4), 465-471.
- Plotnikoff, R.C., Brunet S.B., Kourneya, K.S., Spence, J.C., Birkett, N.J., Marcus, B.M. & Whiteley, J.W. (2007). The efficacy of stage-matched and standard public health materials for promoting physical activity in the workplace: The Physical Activity Workplace Study (PAWS). *American Journal of Health Promotion*, 21(6), 501-509.
- Plotnikoff, R.C., Lippke, S., Courneya, K.S., Karunamuni, N., Sigal, R.J. & Birkett, N. (2007). *Measurement in Physical Education and Exercise Science*, 11(2), 73-91.
- Plotnikoff, R.C., Taylor, L.M., Wilson, P.M., Courneya, K.S., Sigal, R.J., Birket, N., Raine, K. & Svenson, L.W. (2006). Factors associated with physical activity in Canadian adults with diabetes. *Medicine and Science in Sports and Exercise*, 38(8), 1526-1534.
- Plotnikoff, R., Lippke, S., Courneya, K., Birkett, N., & Sigal, R. Physical activity and diabetes: An application of the Theory of Planned Behavior to explain physical activity among type 1 and type 2 diabetes in an adult population sample. *Psychology and Health* (in press).
- Prochaska, J.O. & Velicer, W.F. (1997). The Transtheoretical Model of health behavior change. *American Journal of Health Promotion*, 12, 11-12.
- Prochaska, J.O. & Marcus, B.H. (1994). The Transtheoretical Model: Applications to the cessation of smoking. *Journal of Consulting and Clinical Psychology*, 56, 520-528.
- Prochaska, J.O., Velicer, W.F., Rossi, J.S., Goldstein, M.G., Marcus, B.H., Rakowski, W., Fiore, C., Harlow, L.L., Redding, C.A., Rosenbloom, D. & Rossi, S.R. (1994). Stages of change and decisional balance for twelve problem behaviors. *Health Psychology*, 13, 39-46.
- Renner, B., Spivak, Y., Kwon, S. & Schwarzer, R. (2007). Does age make a difference? Predicting physical activity of South Koreans. *Psychology and Aging*, 22(3), 482-493.

- Rusticus, S.A. & Hubley, A.H. (2006). Measurement invariance of the multidimensional body-self relations questionnaire: Can we compare across age and gender? *Sex Roles*, 55, 827-842.
- Skelly, A.H., Dougherty, M., Gesler, W.M., Soward, A.C., Burns, D. & Arcury, T.A. (2006). African American beliefs about diabetes. *Western Journal of Nursing Research*, 28(1), 9-29.
- Stamatakis, E., Ekelund, U. & Wareham, N.J. (2007). Temporal trends in physical activity in England: The health survey for England 1991 to 2004. *Preventive Medicine*, 45, 416-423.
- Steenkamp, J.B. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.
- Stumvoll, M. Goldstein, R. J. & van Haeften, T.W. (2005). type 2 diabetes: Principles of pathogenesis and therapy. *Lancet*, 365, 1333-1346.
- Sullum, J., Clark, M.M. & King, T.K. (2000). Predictors of exercise relapse in a college population. *Journal of American College Health*, 48, 175-180.
- Suminski, R.R. & Petosa, R. (2006). Web-assisted instruction for changing social cognitive variables related to physical activity. *Journal of American College Health*, 54(4), 219-225.
- Towers, A.J., Flett, R.A. & Seebeck, R.F. (2005). Assessing potential barriers to exercise adoption in middle-aged men: Over-stressed, under-controlled, or just to unwell? *International Journal of Men's Health*, 4(1), 13-27.
- Vandenberg, R.J. & Lance, C.E. (2000). A review and synthesis of measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69.
- Warburton, D.E., Nicol, C.W. & Bredin, S.D. (2006). Health benefits of physical activity: the evidence. *Canadian Medical Association Journal*, 174(6), 801-809.

Appendix A –Wording for the 10-item decisional balance scale items.

Over the next six months:

Pro01t1: Physical activity would help me reduce tension or manage stress.

Pro02t1: I would feel more confident about my health by getting regular physical activity.

Pro03t1: I would sleep better.

Pro04t1: Physical activity would help me have a more positive outlook.

Pro05t1: Physical activity would help me control my weight.

Con01t1: Physical activity would take too much of my time.

Con02t1: I would have less time for my family and friends if I participated in physical activity.

Con03t1: I'd be too tired to get physical activity because of my other daily responsibilities.

Con04t1: I'd worry about looking awkward if others saw me being physically active.

Con05t1: Participation in physical activity would cost too much money.

Notes:

¹ For these analyses, testing between diabetes types was based on self-reported type. Respondents who indicated “unknown”, “no response”, or “both” diabetes types were not included in the diabetes subgroup analysis. As a result, sample sizes reported here differ from those reported by Plotnikoff et al. (2006), wherein additional information available to the investigators (e.g. medical prescriptions) was used to reclassify some “unknown”, “no response” and “both type” cases. All cases in both studies were fully accounted for.

² Although our model evaluation processes and criteria were similar in our calibration and validation study phases, we specifically refer to “covariance model testing” instead of “confirmatory factor analyses” in reference to our re-specification procedures so that we do not mislead readers to imply “confirmatory” testing of the revised model during the calibration phase. The modeling techniques were similar, but re-specification was inherently exploratory.

³ We did not end up relying upon modification indices during the validation process. However we did subjectively confirm both parameter estimates and fit indices were very similar for models when estimated based on the AMOS default missing data option compared to estimates based on imputed data. Where differences existed, they typically occurred at the third decimal digit and were not substantively meaningful.

Table 1: Exploratory factor analysis, using maximum likelihood factor extraction and oblique rotation - pattern matrix loadings for 10-item decisional balance scale (a) specified to extract two factors, and (b) specified to extract all factors with eigenvalues greater than 1.0.

Item	Two-factor Model		Three-factor Model		
	Pros	Cons	Pros	Cons1	Cons2
pro02t1	.841	-.037	.843	.029	-.115
pro01t1	.818	.017	.817	.034	-.025
pro04t1	.805	.000	.805	.002	.000
pro03t1	.760	-.050	.760	-.026	-.041
pro05t1	.589	.048	.594	-.024	.136
con01t1	.018	.777	.001	.812	-.041
con02t1	.062	.775	.045	.747	.044
con03t1	-.021	.740	-.037	.725	.025
con04t1	-.040	.381	.019	-.048	.707
con05t1	.006	.311	-.036	.119	.511

Table 2: Validation samples and longitudinal analysis goodness-of-fit indices. Bold values exceed invariance criteria.

	Chi-sq	df	χ^2_{diff} (df _{diff})	CFI	CFI _{diff}	TLI	RMSEA
Validation Random Samples							
Validation 1	89.338	32	-	.980	-	.971	.048
Validation 2	72.007	32	-	.987	-	.982	.040
Model a (equal form)	161.345	64	-	.983	-	.977	.031
Model b (equal loadings)	172.972	71	11.627 (7)	.983	< .001	.978	.030
Model c (equal factor variances)	177.306	74	15.961 (10)	.982	.001	.979	.030
Model d (equal factor covariances)	186.528	77	25.183 (13)	.981	.002	.978	.030
Model e (equal indicator intercepts)	179.835	78	18.490 (14)	.983	< .001	.980	.029
Model F (equal latent means)	180.403	81	19.058 (17)	.983	< .001	.981	.028
Longitudinal Analysis							
Time 1	98.362	32	-	.989	-	.984	.037
Time 2	92.071	32	-	.990	-	.986	.035
Time 3	143.809	32	-	.981	-	.973	.048
Model a (equal form)	334.242	96	-	.986	-	.981	.023
Model b (equal loadings)	363.324	110	29.324 (14)	.986	< .001	.982	.022
Model c (equal factor variances)	714.069	116	380.069 (20)	.966	.020	.940	.033
Model d (equal factor covariances)	412.669	116	78.669 (20)	.983	.003	.980	.023
Model e (equal indicator intercepts)	430.525	124	96.525 (28)	.983	.003	.981	.023
Model F (equal latent means)	472.550	130	138.55 (34)	.981	.005	.980	.024

Table 3: Subgroup goodness-of-fit indices. Bold values exceed invariance criteria.

	Chi-sq	df	χ^2_{diff} (df _{diff})	CFI	CFI _{diff}	TLI	RMSEA
Sex							
Males	61.090	32	-	.990	-	.986	.034
Females	86.68	32	-	.981	-	.973	.047
Model a (equal form)	147.769	64	-	.986	-	.980	.029
Model b (equal loadings)	150.871	71	3.102 (7)	.986	< .001	.983	.027
Model c (equal factor variances)	159.228	74	11.459 (10)	.985	.001	.982	.027
Model d (equal factor covariances)	166.768	77	18.999 (13)	.985	.001	.982	.027
Model e (equal indicator intercepts)	176.85	78	29.081 (14)	.983	.003	.980	.029
Model F (equal latent means)	236.53	81	88.761 (17)	.973	.013	.970	.035
Diabetes Type							
Type 1	56.596	32	-	.985	-	.979	.041
Type 2	73.63	32	-	.959	-	.985	.035
Model a (equal form)	130.247	64	-	.988	-	.983	.026
Model b (equal loadings)	135.922	71	5.675 (7)	.988	< .001	.985	.025
Model c (equal factor variances)	142.158	74	11.911 (10)	.988	< .001	.985	.025
Model d (equal factor covariances)	146.896	77	16.649 (13)	.987	.001	.985	.025
Model e (equal indicator intercepts)	167.809	78	37.562 (14)	.984	.004	.981	.028
Model F (equal latent means)	191.396	81	61.149 (17)	.980	.008	.978	.030
Education Level							
High school or less	81.464	32	-	.982	-	.974	.048
More than high school	58.397	32	-	.991	-	.988	.031
Model a (equal form)	139.867	64	-	.987	-	.982	.028
Model b (equal loadings)	147.693	71	7.826 (7)	.987	< .001	.983	.027
Model c (equal factor variances)	155.903	74	16.036 (10)	.986	.001	.983	.027
Model d (equal factor covariances)	162.999	77	23.132 (13)	.985	.002	.983	.027
Model e (equal indicator intercepts)	175.937	78	36.070 (14)	.983	.004	.980	.029
Model F (equal latent means)	209.913	81	70.046 (17)	.978	.009	.975	.032
Age							
Less than 50 years	52.849	32	-	.982	-	.968	.040
50 to 70 years	60.397	32	-	.989	-	.982	.039
More than 70 years	52.614	32	-	.988	-	.979	.038
Model a (equal form)	168.356	96	-	.987	-	.982	.022
Model b (equal loadings)	188.213	110	19.857 (14)	.986	.001	.983	.021
Model c (equal factor variances)	204.304	116	35.948 (20)	.985	.002	.982	.022
Model d (equal factor covariances)	223.311	122	54.955 (26)	.982	.005	.981	.023
Model e (equal indicator intercepts)	217.185	124	48.829 (28)	.984	.003	.982	.022
Model F (equal latent means)	349.814	130	181.458 (34)	.962	.025	.960	.033

Table 4: Panel A - Freely estimated latent mean differences compared to youngest age group; Panel B - Latent mean differences compared to youngest age group, with pros and cons2 of middle age group constrained to equal youngest age group; Panel C - Correlations amongst constructs with equality constraints on (1) pros with cons2 for all three age groups (2) pros with cons1 for older two age groups (3) cons1 with cons2 for older two age groups. All correlations in Panel C are statistically significant with $p < .01$.

	50 years or younger	Older than 50 years and younger than 70 years	70 years and older
Panel A			
Pros	0	-.07	-.33
Cons1	0	-.40	-.59
Cons2	0	-.10	-.19
Panel B			
Pros	0	0	-.29
Cons1	0	-.37	-.58
Cons2	0	0	-.12
Panel C			
Pros with Cons1	-.29	-.07	-.07
Pros with Cons2	-.12	-.12	-.12
Cons1 with Cons2	.32	.55	.55

Figure 1: Two-factor decisional balance measurement model.

chi-square = 180.162 df = 34 p = .000
 CFI = .942 TLI = .907 RMSEA = .075
 AIC = 242.162

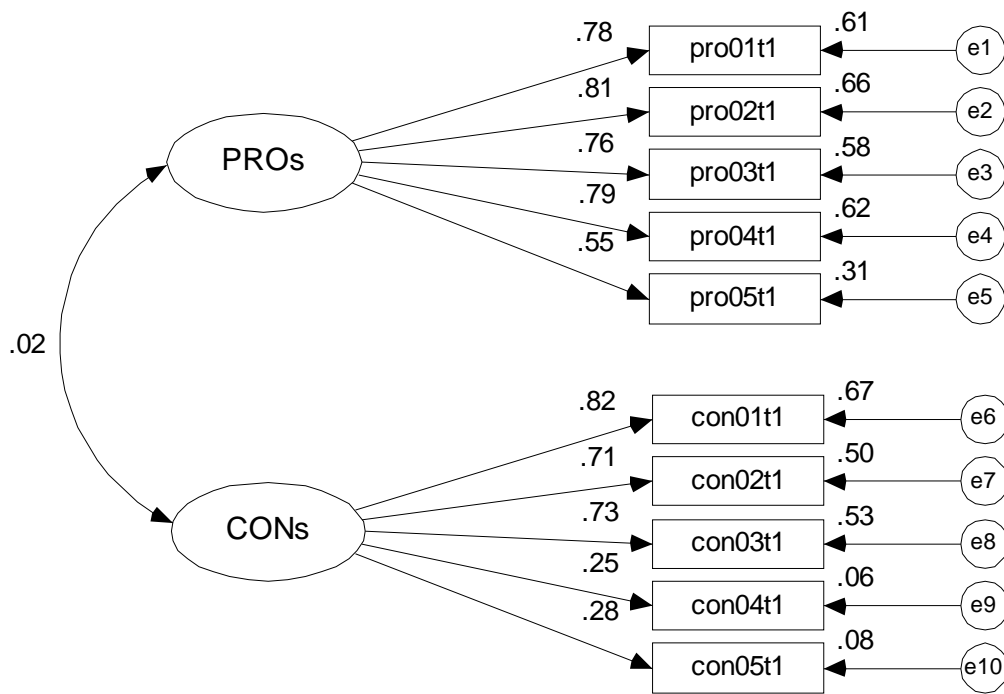


Figure 2: Three-factor decisional balance measurement model.

chi-square = 95.994 df = 32 p = .000
 CFI = .975 TLI = .957 RMSEA = .051
 AIC = 161.994

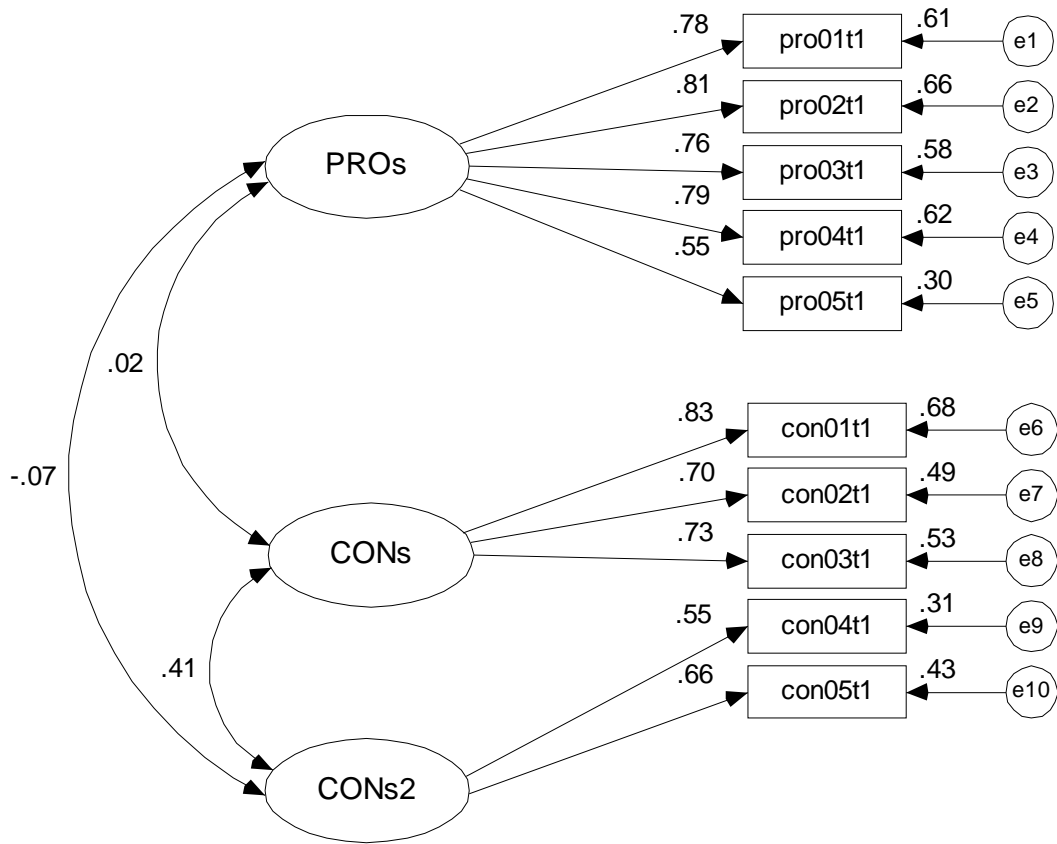


Figure Captions

Figure 1. Standardized parameter estimates and goodness-of-fit indices for two-factor measurement model, fit to calibration sample data. Note: three digit p-value indicated is part of AMOS generated diagram. More appropriate value would be indicated as $p < .001$.

Figure 2. Standardized parameter estimates and goodness-of-fit indices for revised three-factor measurement model, fit to calibration sample data. Note: three digit p-value indicated is part of AMOS generated diagram. More appropriate value would be indicated as $p < .001$.