Work 26 (2006) 429–436
IOS Press

Speaking of Research

Comparing groups on latent variables: A structural equation modeling approach

Dimiter M. Dimitrov
Graduate School of Education, George Mason University, 4400 University Drive, MS 4B3, Fairfax, VA 22030-4444, USA
Tel.: +1 703 993 3842; Fax: +1 703 993 2013; E-mail: ddimitro@gmu.edu

Abstract. Structural equation modeling (SEM) provides a dependable framework for testing differences among groups on latent variables (constructs, factors). The purpose of this article is to illustrate SEM-based testing for group mean differences on latent variables. Related procedures of confirmatory factor analysis and testing for measurement invariance across compared groups are also presented in the context of rehabilitation research.

Keywords: Structural equation modeling, confirmatory factor analysis, latent variable modeling

1. Introduction

The complexity of relationships between variables in rehabilitation settings can be efficiently investigated by the use of multivariate methods of statistical analysis. Testing for group mean differences on a set of observed variables, in particular, typically requires the use of multivariate analysis of variance (MANOVA) or structural equation modeling (SEM). MANOVA is more appropriate when groups are compared on a construct which "emerges" as a linear composite of the observed variables, that is, the observed variables represent causal agents of the construct. An example of such *emergent variable system* is when groups of people categorized by disability type are separated by "stress" – a construct caused by observed dependent variables such as demands of the workplace and relationship with family members. Methodological principles of using MANOVA in the context of rehabilitation research were recently described and illustrated in the "Speaking of research . . ." section of this journal [8].

SEM is more appropriate with a *latent variable system* in which the construct (latent variable) has a causal influence on the observed variables. For example, "self-esteem" is a construct that may underlie the responses of injured workers on specific questionnaire items during a rehabilitation process. An important feature of the SEM methods is that, unlike MANOVA, they provide error-free measures of the latent variables (constructs, factors, subscales) by eliminating the random error of measurement for the observed variables (e.g., questionnaire items) associated with the latent variable(s) [1].

A frequently occurring scenario in rehabilitation research in which SEM can be efficiently employed with a latent variable system is when groups of people are compared on subscales of an instrument (e.g., questionnaire, survey). For example, in a study on prevention of abuse in a workgroup context [15], male and female workers were compared on three subscales of a survey on stressful factors in the workplace: *Hostility* (e.g., "been embarrassed or insulted in front of others"), *Harassment* (e.g., "experienced unwanted sexual advances"), and *Negativity* (e.g., not been praised for your work"). In another study [22], groups of sewing machine operators (e.g., English as a native language versus English as a second language) were compared on three subscales of the Demand-Control Questionnaire

(DCQ [13]): *Demand* (e.g., "my job requires working very fast"), *Control* (e.g., "My job requires that I learn new things"), and *Social support* (e.g., "People I work with are friendly"). In these, and numerous other, examples of rehabilitation research, the questionnaire subscales represent latent variables (factors, constructs) that underlie the persons' responses on questionnaire items. Hereafter, the terms *latent variable*, *construct*, *factor*, and *subscale* will be used interchangeably.

Although SEM provides an excellent framework for the comparison of group means on latent variables, traditional statistical methods such as t-tests, analysis of variance, and analysis of covariance are still predominantly employed in rehabilitation research. In an attempt to help rehabilitation researchers in this direction, this article illustrates SEM methods of comparing groups on latent variables with real data for a population of people with multiple sclerosis. The testing for assumptions required with these methods is also discussed and illustrated.

Table 1
Baseline Model of Two Hypothesized Constructs (Psychosocial Distress and Successful Coping) for People With Multiple Sclerosis

Construct	Items
	In the last month, how often have you . . .
Psychosocial Distress:	
Item 1	been upset because of something that happened unexpectedly?
Item 2	felt that you were unable to control the important things in life?
Item 3	felt nervous and distressed?
Item 4	been angered because of things that happened outside of your control?
Item 5	found that you cannot cope with all the things you had to do
Item 6	felt difficulties were piling up so high that you could not overcome them?
Successful Coping:	
Item 7	felt that things were going your way?
Item 8	dealt successfully with irritating life hassles?
Item 9	felt confident about your ability to handle your personal problems?
Item 10	been able to control irritations in your life?
Item 11	felt that you were effectively coping with important changes that were occurring in you life?

2. Subjects and latent variables

The data were taken from an existing data pool produced with the Employment Preparation Survey Project funded by the National Multiple Sclerosis Society [18]. Two groups of people with multiple sclerosis were formed based on their typical course of illness: (a) *relapsing* – relapsing-remitting or chronic relapsing, and (b) *progressive* – secondary progressive or primary progressive. These two groups, referred to hereafter as relapsing illness group ($n =$ 669) and progressive illness group ($n =$ 359), respectively, were compared on two constructs – *Psychosocial Distress* and *Successful Coping*. The hypothesized model for these two constructs is provided in Table 1 and graphically represented in Fig. 1 [the meaning of the numeric coefficients will be discussed in the next section]. As can be seen, *Psychosocial Distress* is related to six items and *Successful Coping*, to five items of the Employment Preparation Survey. With the sample data, the Cronbach's alpha coefficient of internal consistency reliability for the items associated with *Psychosocial Distress* and *Successful Coping* was 0.88 and 0.81, respectively.

A central question to be addressed before comparing the two groups of people with multiple sclerosis on the constructs of psychosocial distress and successful coping is whether these constructs have the same meaning for each group. First, the data fit of the hypothesized model for the constructs has to be tested with the sample data for each group. This is referred to as testing for "form invariance" of the model across groups [14]. If form invariance is observed, the next step is testing for "measurement invariance" to make sure that the scores on any construct have the same meaning for each of the compared groups. In measurement parlance, the lack of measurement invariance indicates the presence of "differential item functioning" (e.g. [10]). Form invariance and measurement invariance across groups are necessary conditions for meaningful and accurate comparison of groups on construct(s) of interest (e.g. [5,6, 10]). With these two conditions in place, the testing for group mean differences on the construct(s) can be efficiently performed in the framework of SEM (e.g. [3, 9,19]).

3. Testing for form invariance across groups

The validity of the hypothesized model in Table 1 was tested separately for each of the two groups of people with multiple sclerosis – relapsing illness group and progressive illness group. A confirmatory factor analysis (CFA) in the framework of structural equation modeling (SEM) was employed using the computer program for statistical analysis with latent variables Mplus [17]. The CFA results for the two groups are provided jointly in Fig. 1. A number associated with an arrow from a construct to an item is the SEM estimate of the *regression slope* in the linear regression of the

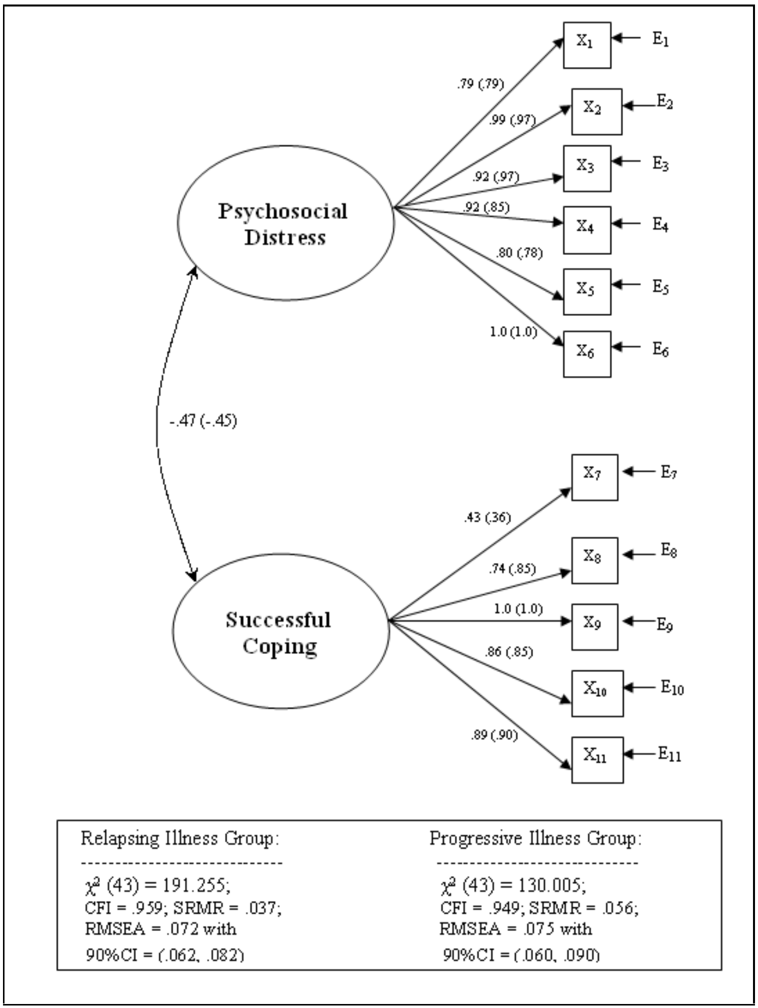

Fig. 1. Baseline model for the constructs of psychosocial distress and successful coping across two groups of people with multiple sclerosis – relapsing illness group and progressive illness group (the regression slopes for this group are given in parentheses). Note. All regression slopes and the correlation between the two constructs (for both groups) are statistically significant at the 0.01 level.

item scores on the (error-free) construct scores (the regression slopes for the progressive illness group are given in parentheses); [the regression slopes are also referred to as *factor loadings* or *structural coefficients*]. In Fig. 1, the regression error terms are denoted E_i; $(i = 1, 2, \ldots, 11)$.

The following four goodness-of-fit indices were used with the CFA in this example – *chi-square fit statistic*, χ^2, *comparative fit index*, CFI, *standardized root mean square residual*, SRMR, and *root mean square error of approximation*, RMSEA, with its 90 percent confidence interval. A relatively good fit is indicated with CFI > 0.90, SRMR < 0.08, and RMSEA < 0.06 (e.g. [4, 11]). Given the sensitivity of the chi-square statistic to sample size, its role in CFA testing for model data fit is more descriptive than inferential. However, this statistic is very useful in a method of comparing the goodness-of-fit for nested models [12]. This method is used in testing for measurement invariance described in the next section. With these clarifications, the values the goodness-of-fit indices reported with Fig. 1 indicate that the hypothesized model fits the data satisfactory well for each of the two groups of people with multiple sclerosis. Therefore, the assumption of form variance across the two groups is met.

Some additional clarifications are necessary to better understand the meaning of form invariance in general. The most parsimonious, yet substantively most meaningful and best fitting model to the data for a group is referred to as *baseline model* for this group (e.g. [6,14]).

In our example, the two groups of people with multiple sclerosis have the same baseline model – the model in Fig. 1, with correlational relationship (two-way arrow) between the two constructs [correlation of -0.47 for the relapsing illness group and -0.45, for the progressive illness group]. In general, the baseline models are not necessarily identical across groups. It may happen, for example, that some constructs are correlated with the baseline model for one group, but not with the baseline model for another group [6]. As shown in previous research [20], only the comparable parameters within the same construct need to be equated across the compared groups before testing for group mean differences on the construct.

4. Testing for measurement invariance across groups

Since it was found that form invariance is in place, we can proceed with testing for "measurement invariance" to determine whether the scores on each construct have the same meaning for each group. As alluded to earlier, a lack of measurement invariance would indicate the presence of differential item functioning thus threatening the validity of results related to group mean differences on the construct (e.g. [6,14,16]). In general, the testing for measurement invariance includes testing for invariance of (a) *regression slopes* – in Fig. 1, the estimates of regression slopes are associated with the one-way arrows from the constructs to the observed variables, X_i, (b) *regression intercepts* (not depicted in Fig. 1), and (c) *error variances* – in Fig. 1, the variances of regression error terms, E_i, $(i = 1, \ldots 11)$. For the purpose of group mean comparison, it is practically sufficient to test only for invariance of the regression slopes and intercepts [21]. As noted in previous research, testing for the invariance of the error variances provides an overly restrictive test of the data [2].

Given the purpose of this article, only the invariance of regression slopes and intercepts across the two groups is tested for the model in Fig. 1. This is done by using the chi-square square test for the difference between two nested models: a model with "invariance assumed" (χ^2_{INVAR}) and model with "no invariance assumed" ($\chi^2_{\mathrm{NO\_INVAR}}$). The invariance of parameters being tested is confirmed when the chi-square difference, $\chi^2_{\mathrm{DIFF}} = \chi^2_{\mathrm{INVAR}} - \chi^2_{\mathrm{NO\_INVAR}}$, is not statistically significant (e.g. [17]). Specifically, the following three-step testing procedure is applied:

Table 2
Tests for the Invariance of Regression Slopes and Intercepts Across Two Groups of People with Multiple Sclerosis (Relapsing Illness and Progressive Illness)

Model	χ^2	*df*	$\Delta\chi^2$	Δdf
Model 0	321.26	86		
Model 1	327.63	95	6.63	9
Model 2	347.43	104	19.80*	9
Model 2P	338.39	103	10.76	8

Note. Model 0: Non invariant slopes and intercepts;
Model 1: Invariant slopes, non invariant intercepts
Model 2: Invariant slopes and invariant intercepts;
Model 2P: Invariant slopes and invariant intercepts, with a "free" intercept for Item 2 (partial invariance).
$^*p < 0.05$.

- Model 0: The model in Fig. 1 is fit in the two groups together allowing all parameters, including regression slopes and intercepts, to be free – that is, no invariance of parameters across the two groups is assumed.
- Model 1: The model in Fig. 1 is fit in the two groups together, with the regression slopes held equal across the groups. Since Model 1 is nested within Model 0, the chi-square difference for the two models is used to test for invariance of the regression slopes.
- Model 2: The model in Fig. 1 is fit in the two groups together, with both regression slopes and intercepts held equal across the groups. Since Model 2 is nested within Model 1, the chi-square difference for the two models is used to test for invariance of the regression intercepts.

Table 2 provides the testing results for measurement invariance obtained with the computer program Mplus [17]. The Mplus syntax code (input) for the testing models is given in Fig. 2. As the results in Table 2 show, the chi-square difference for Model 0 versus Model 1 ($\Delta\chi^2 = 6.63$, $df = 9$) is not statistically significant thus providing evidence for the invariance of the regression slopes across the two groups. Further, the chi-square difference for Model 1 versus Model 2 ($\Delta\chi^2 = 19.80$, $df = 9$) is statistically significant at the 0.05 level (but not at the 0.01 level). So, there is no perfect invariance of the intercepts across the two groups, but neither is there evidence of complete inequality. This situation is termed *partial measurement invariance* [6,14]. As previous studies show, given the stringent nature of the hypotheses for invariance, the invariance is a matter of degree estimated by the proportion of parameters that are invariant (e.g. [3,6,14, 16]).

To determine the degree of partial measurement invariance in our case, Model 2 has to be modified by

```
MODEL 0:
TITLE:  Model 0 [without measurement invariance]
DATA:  FILE IS myfile.dat;
VARIABLE:  NAMES ARE   group  X1-X11;
GROUPING  IS  group  (0 = relapse, 1 = progress);
USEVARIABLES ARE  X1-X11;
ANALYSIS:  TYPE IS GENERAL;
OUTPUT:  STANDARDIZED;

MODEL: DISTRESS BY X1-X6;
            COPING BY X7-X11;
            DISTRESS WITH COPING;
MODEL progress:  DISTRESS BY X2-X6;
                 COPING BY X8-X11;
                 DISTRESS WITH COPING;
------------------------------------------------------------
MODEL 1:
TITLE:  Model 1 [invariant slopes, free intercepts]
DATA:  FILE IS myfile.dat;
VARIABLE:  NAMES ARE   group  X1-X11;
GROUPING  IS  group  (0 = relapse, 1 = progress);
USEVARIABLES ARE   X1-X11;
ANALYSIS:  TYPE IS GENERAL;
OUTPUT:  STANDARDIZED MODINDICES (3.84);

MODEL:  DISTRESS BY X1-X6;
             COPING BY X7-X11;
             DISTRESS WITH COPING;
----------------------------------------------------------------------------
MODEL 2: Same as MODEL 1, with two  changes:

TITLE: Model 2 [invariant slopes and intercepts];
ANALYSIS:  TYPE = MEANSTRUCTURE;
-------------------------------------------------------------
MODEL 2P: Same as MODEL 2, but add the following
statement in the model part to "free" the intercept for X2:

MODEL progress: [X2];
```

Fig. 2. Mplus input for testing measurement invariance with the mean structures method of comparing group means on two constructs – DISTRESS (psychosocial distress) and COPING (successful coping), for two groups of people with multiple sclerosis – "relapse" (relapsing illness group) and "progress" (progressing illness group).

setting some intercepts "free" (non invariant) across the two groups. Which intercepts to start freeing depends on the values of their modification indices reported with Mplus for Model 2. In general, the MI value for a parameter gives the expected drop in the model chi-square value if this parameter is freely estimated [7,12, 17]. The MI is statistically significant if it exceeds 3.84 (the chi-square value with *df* = 1). With the Mplus results for Model 2, the MIs for the intercepts of three items were statistically significant: Item 1 (MI

= 6.18), Item 2 (MI = 9.02), and Item 3 (MI = 4.99). The intercept for the item with the largest MI (Item 2) was set "free" in Model 2 thus obtaining Model 2P. With this, the chi-square value for Model 2 (347.43) dropped to 338.39 for Model 2P. As shown in Table 2, the chi-square difference for Model 1 versus Model 2P ($\Delta\chi^2 = 10.76$, $df = 8$) is not statistically significant. Therefore, there is no need to free additional intercepts. In general, there is no hard rule as to what degree of partial invariance is acceptable; it is up to researchers to decide, as long as the observed degree of invariance is reported with the results (e.g. [6,14]). In our case, with 10 (out of 11) invariant intercepts, the degree of partial measurement invariance across the two groups seems sufficiently high to proceed with testing their mean difference on the two constructs, psychosocial distress and successful coping.

5. Comparing group means on latent variables

With the assumption of measurement invariance across groups met, the door is open for comparison of the group means on latent variables of interest. There are two major approaches to addressing this task in the context of SEM – *structured means analysis* and *group code analysis* (e.g. [1,2,9,17]). These two methods are briefly discussed and illustrated here in the context of comparing the two groups of people with multiple sclerosis (relapsing illness/progressive illness) on psychosocial distress and successful coping.

5.1. Structured means analysis

Structured means analysis is applied with models that contain intercepts for the purpose of estimating group means on a construct. Under the analytic model of structured means, the mean of each group on an observed variable, X_k, is obtained from its linear regression on the construct mean for this group: Mean $(X_k) = \tau_k + \lambda_k$[Mean($\xi$)], where ξ is the construct, λ_k is the regression slope, and τ_k is the regression intercept. To estimate the difference between two group means on the construct, one of the groups is chosen to serve as a reference group and its mean on the construct is fixed to zero. With this, the construct mean of the other group represents the difference between the construct means of the two groups. This is derived from the analytic model of structured means under the assumption that λ_k and τ_k do not change in value across the two groups (e.g. [3,9]). The testing for this assumption (invariance of the regression slopes and intercepts) was conducted in the previous sections.

For the two groups of people with multiple sclerosis, the structured means analysis was performed using the Mplus input for Model 2P (see Fig. 2). The relapsing illness group was chosen as a reference group – in the Mplus input this is reflected with the coding (0 = relapse, 1 = progress) in the grouping syntax line. By doing so, the difference between the two group means on each construct equals the mean of the non-reference group (progressive illness) on the construct. Specifically, the mean of the progressive illness group was -0.01 (not statistically significant) on psychosocial distress and -0.12 (statistically significant) on successful coping. Therefore, the conclusion is that the two groups do not differ on psychosocial distress, but they differ on successful coping, with the higher mean score for the reference group since the group difference is negative in sign (-0.12). In addition, since the intercept for Item 2 was "freed" across the two groups with the model used for this mean structure analysis (Model 2P), the two groups may differ on this particular item. More information on this is provided with the illustration of the other SEM method (group code analysis) for group mean differences on a construct.

5.2. Group code analysis

While the structured means approach to group mean differences keeps the data from the two groups separate (like a t-test), the group code analysis uses the data from both groups in a single SEM model (like dummy coding in a linear regression model). Also, while the invariance of slopes and intercepts is a prerequisite for testing group mean differences with the structured means method, the group coding method requires that the (single) measurement model holds in both groups thus including invariance of the slopes, construct variance, and error variances/covariances. However, if the invariance with the structured means analysis holds for all parameters in the measurement models for the two groups, the structured means analysis and the group code analysis yield identical outcomes with regard to group mean differences (e.g. [1,17]). A combination of the two methods can be particularly useful with more complex models of group mean differences – this is the case, for example, when the group means are adjusted for pretest differences on the construct(s) of interest (e.g. [1]).

The group code analog to the mean structure analysis conducted with Model 2P is represented with the

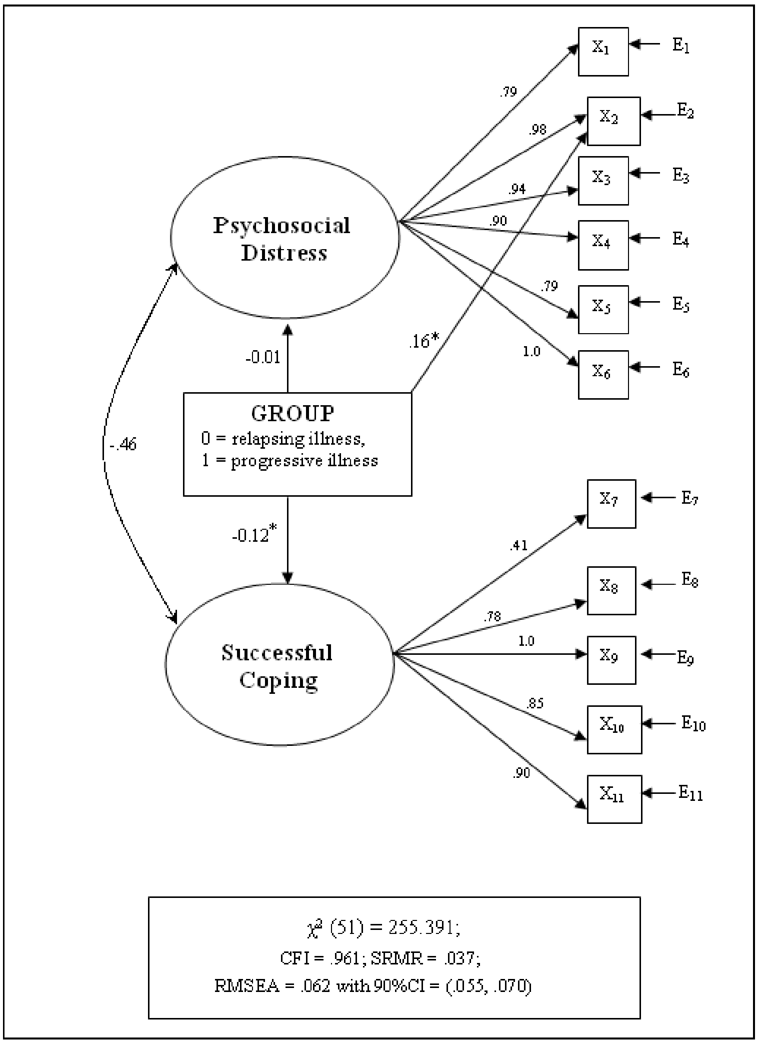

Fig. 3. Group code model for testing the mean difference between two groups of people with multiple sclerosis on constructs of psychosocial distress and successful coping. Note. All regression slopes and the correlation between the two constructs are statistically significant at the 0.01 level. $^*p < 0.05$.

SEM model in Fig. 3. The one-way arrow for "direct effect" of the grouping variable (GROUP) to the observed variable X_2 is the group code analog to "freeing" the intercept for Item 2 with the mean structured approach. Without going into testing for invariance beyond the equality of regression slopes, which was previously confirmed, let us examine the results obtained with the model in Fig. 3 using Mplus.

First, the values of the goodness-of-fit indices CFI, SRMR, and RMSEA indicate an acceptable data fit for the group code model (see, Fig. 3). Note also that the regression slopes associated with the one-way arrows from the two constructs to the observed variables are almost identical to their counterparts (for each group separately) in Fig. 1. The same holds for the correlation between the two constructs. Further, the results for group mean differences on each construct are identical to those obtained earlier with the mean structures method. Indeed, the structural coefficient associated with the one-way arrow from GROUP to Psychosocial Distress, representing the group mean difference on this construct, equals -0.01 (not statistically significant). Also, the structural coefficient with the one-way arrow from GROUP to Successful Coping is statistically significant and equals -0.12. Clearly, the group code method and mean structures method yield identical outcomes with regard to the group mean differences on psychosocial distress and successful coping. In Fig. 3, the one-way arrow for direct effect from GROUP to the observed variable X_2 is the group code analog to freeing the intercept for Item 2 across the two groups with the mean structures method (Model 2P). Since the estimate of this direct effect (0.16) is positive and statistically significant, given the coding of the two groups (0 = relapsing, 1 = progressive), we conclude that the progressive illness group performed better than expected on the observed variable X_2 (see, Item 2 in Table 1).

6. Conclusion

Structural equation modeling (SEM) provides a versatile analytic framework for testing group mean differences on latent variables (constructs). A frequently occurring situation is rehabilitation research when SEM-based testing of group mean differences is suitable (yet, still not sufficiently utilized) is when groups are compared on the subscales of an instrument (e.g., questionnaire or survey). The hope is that the SEM methods (and testing for their assumptions) described and illustrated in this article will provide *Work* readers with important methodological principles and "Know-how" in comparing groups on complex constructs in the context of rehabilitation research and/or assessment.

References

[1] L.S. Aiken, J.A. Stein and P.M. Bentler, Structural equation analyses of clinical subpopulation differences and comparative treatment outcomes: Characterizing the daily lives of drug addicts, *Journal of Consulting and Clinical Psychology* **62** (1994), 488–499.

[2] P.M. Bentler, *EQS: Structural Equations Program Manual*, Multivariate Software, Encino, CA, 2004.

[3] K.A. Bollen, *Structural Equations with Latent Variables*, John Wiley, NY, 1989.

[4] M.W. Browne and R. Cudeck, Alternative ways of assessing model fit, in: *Testing Structural Equation Models*, K.A. Bollen and J.S. Long, eds, Sage, Newbury Park, CA, 1993, pp. 136–162.

[5] B.M. Byrne, Testing for multigroup invariance using AMOS graphics: A road less traveled, *Structural Equation Modeling* **11** (2004), 272–300.

[6] B.M. Byrne, R.J. Shavelson and B. Muthén, Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance, *Psychological Bulletin* **105** (1989), 456–466.

[7] B.M. Byrne, *Structural Equation Modeling with AMOS: Basic Concepts, Applications and Programming*, Lawrence Erlbaum, Mahwah, NJ, 2001.

[8] D.M. Dimitrov and P.D. Rumrill, Multivariate methods in rehabilitation, *WORK* **24** (2005), 205–212.

[9] G.R. Hancock, Structural equation modeling methods of hypothesis testing of latent variable means, *Measurement and Evaluation in Counseling and Development* **30** (1997), 91–105.

[10] P.W. Holland and H. Wainer, *Differential Item Functioning*, Erlbaum, Hillsdale, NJ, 1993.

[11] L. Hu and P.M. Bentler, Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives, *Structural Equation Modeling* **6** (1999), 1–55.

[12] K.G. Jöreskog and D. Sörbom, *LISREL's Reference Guide*, Scientific Software, Inc, Chicago, 1989.

[13] R. Karasek and T. Theorell, *Healthy Work, Stress, Productivity and the Reconstruction of Working Life*, Basic Books, New York, 1990.

[14] D.W. Levine, R.M. Kaplan, D.F. Kripke, D.J. Bowen, M.J. Naughton and S.A. Shumaker, Factor structure and measurement invariance of the women's health initiative insomnia rating scale, *Psychological Assessment* **15** (2003), 123–136.

[15] L.L. Mangione and T.W. Mangione, Workgroup context and the experience of abuse: An opportunity for prevention, *WORK* **16** (2001), 259–267.

[16] W. Meredith, Measurement invariance, factor analysis and factorial invariance, *Psychometrika* **58** (1993), 525–543.

[17] L.K. Muthén and B.O. Muthén, *Mplus: Statistical Analysis with Latent Variables, User's Guide*, Muthén & Muthén, Los Angeles, CA, 2003.

[18] R. Roessler, P. Rumrill and M. Hennessey, *Employment Concerns of People with Multiple Sclerosis: Building a National Employment Agenda*, National Multiple Sclerosis Society, New York.

[19] K.H. Sousa and F.F. Chen, Latent mean structure analysis: A technique to compare group differences in a second-order quality of life model, *Journal of Nursing Measurement* **11** (2003), 253–266.

[20] C.E. Werts, D.A. Rock, R.L. Linn and K.G. Jöreskog, Comparison of correlations, variances, covariances, and regression weights with and without measurement error, *Psychological Bulletin* **83**, 1007–1013.

[21] K.F. Widman and S.P. Reise, Exploring the measurement invariance of psychological instruments: Applications in the substance use domain, in: *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research*, K.J. Bryant, M. Windle and S.G. West, eds, American Psychological Association, Washington, D.C., 1997, pp. 281–324.

[22] R.M. Williams, G. Sundelin and M.L. Schmuck, Reliability of the demand-control questionnaire for sewing machine operators, *WORK* **16** (2001), 71–75.