**Qing Shu - BUS-Z 798 - Assignment 3**

**Part 1 - CFA Analysis**

1. What are the distinctions between exploratory and confirmatory factor analysis?

* EFA doesn’t assume a measurement model while CFA must assume.
* The loadings of EFA mirror the true score of CFA. We may only need to concern the absolute value of the loadings (to evaluate item performance), while both the direction and the value of true score should be paid attention to, because it represents the relationships between measures and constructs.
* The error variance of CFA mirrors the uniqueness of EFA. The uniqueness is something left (after factor extraction). It represents information that are not shared with other measures. The error variance is, however, important for model fit in CFA. It has theoretical meaning, i.e., the random effect in the relationship of measures and constructs, thus, must be included in the model.
* In EFA, unidimensionality is not assumed a priori. Items can actually correlate with all factors. In contrast, CFA requires the researcher to specify the association between items and constructs in advance, and one item should only be specified with a single factor.
* The result 'model' in EFA are the most likely model statistically but it has no a prior theoretical meaning. Meanwhile, in CFA, particular through a two-step modeling approach, the model is first tested under a theoretical setting and then compared with other nested models to evaluate its statistical fitness.

1. There are various models relating constructs to measures and criteria for choosing among those models. What are the implications of these models for developing and evaluating measures in your area?

First, all these models together encourage me to reflect the single-measure operationalization in my research area – innovation strategy. Before reading the work of Edwards and Bagozzi (2000), I didn’t realized that the relationships between constructs and measures can be so diversified and sophisticated. Innovation research through secondary data seldom implements questionnaires. Scholars tend to build measures through constructing ratios or index based on, for instance, patent data. However, there are few researchers asking whether such ratios are the reflective or formative measures of the target constructs. It might be due to that a single-measure operationalization often implies no better alternatives for measurement. However, such questions still deserve attention, particular in helping to articulate the underlying mechanism between constructs.

Second, in the field of innovation research, patent data are often a reflective measure, but the work of Edwards and Bagozzi (2000) inspires me to think whether patent data can be formative. Most often, patent data are constructed as reflective measures because they are the result of the firms’ behavior (e.g., exploratory innovation, open innovation) or capability (e.g., technology depth or breath). However, I think patent data can be formative in some context. If patents are used to build a construct which measures the technological barriers a firm can set on against its competitors, then patent data in this condition would constitute a formative measure. Because the firm’s capability to prevent competitors’ entry would be a result of its legal ownership of some key patents over some technologies.

1. Using last week's EFA dataset loaded to canvas, conduct 2 CFA analyses using the sem command in STATA. Both models should be basic with 2 latent factors (engagement and job satisfaction). However, the difference will be the items used to measure the latent factors. Model 1 should use Y23, Y28, and Y36 for engagement and Y10, Y11, and Y12 for job satisfaction. Model 2 should use Y28, Y31, and Y32 for engagement and Y11, Y12, and Y13 for job satisfaction.

**Table 1 Reports of fit statistic**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Model 1 | Model 2 | Suggested by Nye (2023) |
| chi-square (P-value) | 65.551(0.000) | 6.491(0.592) | Smaller is better |
| degrees of freedom | 8 | 8 | - |
| RMSEA | 0.112 | 0.000 | ≤.06 |
| CFI | 0.877 | 1.000 | ≥.95 |
| SRMR | 0.080 | 0.016 | ≤.06 |

As shown I Table 1, Model 2 obviously fit better. In both of models, the indicators are significantly correlated to their respective factors (p-value = .000). However, for the Model 1, the modification indices show that fitness can be largely improved if some constrains are loosened. This may imply that some items are designed poorly so that they reflect information of other factors, such as Y23, Y36 and Y10 (Table 2).

**Table 2 The Modification indices of Model 1**

| **Modifiication** | **MI** | **df** | **P>MI** | **EPC** | **EPC (Standard)** |
| --- | --- | --- | --- | --- | --- |
| Y23 <- JobSatisfication | 16.507 | 1 | 0.00 | 0.7073489 | 0.2225181 |
| Y36 <- JobSatisfication | 4.261 | 1 | 0.04 | -0.4023798 | -0.1333066 |
| Y10 <- Engagement | 44.685 | 1 | 0.00 | 1.142021 | 0.340202 |
| Y12 <- Engagement | 8.469 | 1 | 0.00 | -0.5787719 | -0.1763426 |
| cov(e.Y23,e.Y28) | 4.261 | 1 | 0.04 | -0.1494548 | -0.2627405 |
| cov(e.Y23,e.Y11) | 3.994 | 1 | 0.05 | 0.1109947 | 0.0993395 |
| cov(e.Y28,e.Y36) | 16.508 | 1 | 0.00 | 1.392856 | 3.166598 |
| cov(e.Y28,e.Y10) | 23.359 | 1 | 0.00 | 0.1469539 | 0.3151433 |
| cov(e.Y10,e.Y11) | 8.469 | 1 | 0.00 | -0.2664749 | -0.2909281 |
| cov(e.Y11,e.Y12) | 44.685 | 1 | 0.00 | 1.735621 | 2.473707 |

1. The third model you run should be on the same items as model 2, but instead of running a 2 factor CFA with engagement and job satisfaction, I want you to **have all the items load onto a single factor**.

**Table 3 Reports of fit statistic of the single-factor model**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Model 2 | Single-factor model | Suggested by Nye (2023) |
| chi-square (P-value) | 6.491(0.592) | 573.559(0.000) | Smaller is better |
| degrees of freedom | 8 | 9 | - |
| RMSEA | 0.000 | 0.329 | ≤.06 |
| CFI | 1.000 | 0.575 | ≥.95 |
| SRMR | 0.016 | 0.207 | ≤.06 |

This result (Table 3) shows that single-factor model is not appropriate.

**Part 2 - SEM Analysis**

1. Using the SEM dataset uploaded to Canvas, run two SEM models with at least 2 exogenous latent variables, at least one latent mediator variable, and at least one latent outcome variable (all of your choosing).

I have problem in visualized the SEM graphic.

For the partially mediated Structural model (P Model), I set:

(Performance <- GOA TaskConf CM) (GOA <- CM)

For the fully mediated Structural model (F Model), I set:

(Performance <- GOA TaskConf) (GOA <- CM)

Below are the results.

|  |  |  |  |
| --- | --- | --- | --- |
|  | P Model | F Model | Suggested by Nye (2023) |
| chi-square (P-value) | 10529.020 (0.000) | 10534.374 (0.000) | Smaller is better |
| degrees of freedom | 204 | 231 | - |
| RMSEA | 0.393 | 0.393 | ≤.06 |
| CFI | 0.345 | 0.345 | ≥.95 |
| SRMR | 0.143 | 0.146 | ≤.06 |

Inappropriate fitness here implies no model can be retained.

1. Using the same dataset and variables as in question 1, test the hypothesis that the proposed mediator(s) mediate(s) the relation between the independent variables and the outcome variable. To test this hypothesis, use the steps described in the link above to compute bias-corrected 95% confidence intervals around the indirect effect, in addition to the relevant model statistics. Report the results of the relevant analyses you conducted to arrive at your conclusion.

I try to make this function work, but have no idea about how to set the number parameters in the el().

program indireff, rclass

sem (GOA -> goa\*) (TaskConf -> conf\*) (CM -> cong\*) ///

(Performance -> TInrole\*) ///

(Performance <- GOA TaskConf CM) (GOA <- CM) ///

, latent(GOA TaskConf CM Performance)

estat teffects

mat bi = r(indirect)

mat bd = r(direct)

mat bt = r(total)

return scalar indir = el(bi,1,3)

return scalar direct = el(bd,1,3)

return scalar total = el(bt,1,3)

end

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155-174. <https://doi.org/10.1037//1082-989x.5.2.155>

Nye, C. D. (2023). Reviewer resources: Confirmatory factor analysis. *Organizational Research Methods*, *26*(4), 608-628. <https://doi.org/10.1177/10944281221120541>