



A CROSS – NATIONAL RESEARCH OF YOUTH OFFENDING: SUGGESTING SOLUTIONS FOR METHODOLOGICAL CHALLENGES

Social Science

Myunghoon Roh

Department of Social Science and Humanities, Texas A&M University – San Antonio, One, University Way, San Antonio, TX 78224

KEYWORDS

INTRODUCTION

Contemporary criminology theoretical frameworks, as well as the metrics for evaluating them, were developed and tested mainly within Western cultural settings, especially the United States (Karstedt, 2001). Gottfredson and Hirschi's (1990) self-control theory and Grasmick et al's (1993) self-control scale are a case in point. Fields outside of criminology have more readily embraced transnational and comparative methods. For instance, the cross-cultural approach is drawn originally from the field of sociology and cultural anthropology. Durkheim's statement illustrates this point: "*Comparative sociology is not a particular branch of sociology; it is sociology itself*" (Durkheim, 1938, 1951: 139).

The recent globalization of political, social, economic, and technological issues as well as the increasing recognition that crime is a worldwide phenomenon have led criminologists to focus more on the cross-cultural approach. According to Bennett's (2004) presidential address, the events of September 11th dramatically increased interest in comparative criminology and criminal justice. In addition, Dammer, Riechel, & He (2005) suggested several factors which escalated interest in the cross-cultural approach: European Union's eastward expansion, the modernization of economy, opening of previously restricted-borders, and widespread transcontinental mobility.

The goal of comparative criminology is to test the scientific generalizability of the causal mechanism purposed by criminology theories and conceptualizations (Miller, Jennings, Alvarez-Rivera, & Lanza-Kaduce, 2009; Botchkovar, Tittle, & Antonaccio, 2009). Testing theoretical concepts through a cross-cultural approach allows criminologists to compare historical, political, social, economic, and cultural influences on crime across different countries (Vazsonyi, Pickering, Junger, & Hessing, 2001; Meneses & Akers, 2011). However, the mostly Western origin of contemporary criminological perspectives and scales has placed comparative criminologists in a difficult position.

Comparative criminologists need to consider the methodological challenges for cross-cultural analyses in the field of criminology. It is possible that due to the origin of theoretical frameworks from a single cultural setting (i.e. United States) the same set of questions in the survey might have different meanings in different cultures or countries. Thus, the validity of the results from cross-cultural studies without ensuring for invariance of measurements is weak (Rodriguez, Perez-Santiago, & Birkbeck, 2015). Therefore, measurement invariance testing is a critical issue for meaningful comparison across different countries. When researchers are comparing groups across different countries, it needs to be ascertained whether they are measuring the same psychometric constructs in all groups in the same manner or not. Unfortunately, the issues of measurement invariance have largely been neglected in previous cross-cultural empirical criminology research. However, measurement invariance testing has received increasing attention in and scholars have begun to overcome the methodological challenges cross-cultural research entails (e.g. Ren, Zhao, He, Marshall, Zhang, & Zhao, 2015).

Therefore, the purpose of current study is to address methodological challenges of cross – national criminology research. To achieve this goal, I conduct multiple – group confirmatory factor analyses in order to test the cross – cultural comparability of measurement (Farrall et al., 1999; Farrall, 2004). Ultimately, this study will answer the following question; Does measurement invariance testing increase the validity of

the result of cross – national research?

RESEARCH METHODOLOGY

The second wave of International Self-Report Delinquency Study (ISR2) was utilized for the current study, which examines two different types of delinquent offending from 59,699 adolescents in 25 countries. The study includes two Anglo-Saxon countries (USA and Ireland), four Mediterranean countries (Spain, Italy, Cyprus, and Portugal), five northern European countries (Finland, Sweden, Norway, Denmark, and Iceland), six Western European countries (Belgium, France, Switzerland, Germany, Austria, and Netherlands), and eight post-Socialist countries in Europe (Russia, Bosnia and Herzegovina, Slovenia, Hungary, Czech Republic, Poland, Lithuania, and Estonia). Although the ISR2 study contains data collected from non-European countries, author used only European countries plus United States to increase comparability.

Measurement

Predictors from Social Bonding Theory

Attachment to parents was measured by the two items measuring closeness to mother and father. The respondents in the ISR2-2 were asked to answer two items on a 4-point Likert-type scale ranging from 1 (not at all) to 4 (very well) that asked, "How do you usually get along with the man you live with (father, stepfather)?" and "How do you usually get along with the woman you live with (mother, stepmother)?" Results from principal component analysis report that these two items are correlated in the same underlying structure with a Cronbach's alpha level of .62. The concept of **attachment to school** was measured by three items on a 4-point Likert-type scale ranging from 1 (not at all true) to 4 (very true) that asked, "If I had to move I would miss my school," "Teachers do notice when I am doing well and let me know," and "I like my school." Results from principal component analysis found that these three items are correlated in the same underlying structure with a Cronbach's alpha level of .64. **Attachment to neighborhood** was measured by three items that asked, "If I had to move, I would miss people in neighborhood," "My neighbors notice when I am misbehaving and let me know," and "I like my neighborhood." Items ranged from 1 (not at all true) to 4 (very true). Results from principal component analysis report that these three items are correlated in the same underlying structure with a Cronbach's alpha level of .61.

Predictors from Lifestyle/Routine Activity Theory

According to Wikstrom (2006), individuals' acts of violence are an outcome of the causal interaction between their propensity to engage in crime and their exposure to situational factors to engage in acts of violence. Thus, following this assumption, the author operationalized association with deviant peers in the ISR2-2 as a concept of exposure. Respondents in ISR2-2 were asked how many friend they have who have done the following delinquent acts: "used soft or hard drugs like weed, hash, XTC, speed, heroin or coke," "Stole something from a shop or department store," "Entered a building with the purpose to steal something," "Threatened somebody with weapon or to beat him up, just to get money or something from him," and "Beat someone up or hurt someone badly with something like a stick or a knife." Results from principal component analysis found that these five items are correlated in the same underlying structure with a Cronbach's alpha level of .70. Most of the sample did not have friends who had done such things (59.25%). However, almost 20% (19.46%) had friends who had engaged in one of these activities and close to 12% had friends that had engaged in two of these activities (11.71%). Only a small group had genuinely delinquent friends who had engaged in three or more illegal

activities (9.59%).

Predictors from Self-Control Theory

As a modified Grasmick et al. (1993) scale, the ISRD-2 contains 12 items covering **impulsivity**, **risk-seeking**, **self-centeredness**, and **temperament**. Each of the 12 item ranges from 1 (strongly agree) to 4 (strongly disagree). Given that the current dissertation is focusing on individuals' level of self-control, the summation of all 12 items represent their level of self-control. The first three items are a measure of **impulsivity**. They asked respondents how strongly they agree or disagree with the statements: "I act on the spur of the moment without stopping to think," "I do whatever brings me pleasure here and now, even at the cost of some distant goal," and "I am more concerned with what happens to me in the short run than in the long run." Results from principal component analysis reported as; these three items are correlated in the same underlying structure with a Cronbach's alpha level of 0.57. The second three items are a measure of **risk-seeking**. They asked respondents how strongly they agree or disagree with the statements: "I like to test myself every now and then by doing something a little risky," "Sometimes I will take a risk just for the fun of it," and "Excitement and adventure are more important to me than security." Results from principal component analysis found that these three items are correlated in the same underlying structure with a Cronbach's alpha level of .79. The third three items are a measure of **self-centeredness**. They ask respondents how strongly they agree or disagree with the statements: "I try to look out for myself first, even if it means making things difficult for other people," "If things I do upset people, it's their problem not mine," and "I will try to get the things I want even when I know it's causing problems for other people." Results from principal component analysis found that these three items are correlated in the same underlying structure with a Cronbach's alpha level of .68. The last three items are a measure of **temperament** by asking respondents to rate the statements: "I lose my temper pretty easily," "When I am really angry, other people better stay away from me," and "When I have a serious disagreement with someone, it's usually hard for me to talk calmly about it without getting upset." Results from principal component analysis found that these three items are correlated in the same underlying structure with a Cronbach's alpha level of .69.

Analytic Strategy

Three different analyses are conducted to accomplish the goals of the current study. First, the factorial structure of the measurement model is assessed by examining the baseline model of the multigroup confirmatory factor analysis on the sample of each participating countries in the ISRD-2 study respectively. Then, author tests measurement invariance across countries using a multigroup confirmatory factor analysis. Finally, if measurement invariance across different countries was seen, then the research hypotheses of the current study are tested.

Factorial Structure Test (Single – Country Analyses)

The factorial structure of the proposed measurement model, the baseline model of the multigroup confirmatory factor analysis model in each participating country of the ISRD-2 study is tested. The underlying assumption of the factorial structure test is that respondents' responses on the measurement items have the same factor loading across different countries as well as that each item has similar variance when the test is conducted separately for each country.

Multi-Group Confirmatory Factor Analysis (MGCEFA)

The multiple group confirmatory factor analysis to answer the second research question on methodological challenges of comparative research was employed: Are the theoretical instruments of interest invariant across different countries? The multigroup CFA using Structural Equation Modeling techniques to answer the research question with Maximum Likelihood estimation was tested. The model and instruments are then evaluated by model fit indices using the recommended criteria indicated below.

Configural Invariance Test

Author used a configural invariance test to examine whether the proposed five-factor model is invariant across different participating countries in the ISRD-2 study. Author tested whether the countries have the same factor structure. Specifically, author examined whether the fixed and free factor loadings have the same pattern in the configural invariance model for all participating countries. Using the

goodness-of-fit index criteria, author evaluated the model fit of the configural invariance. As a baseline model, this configural model is compared to the other model with more constraints.

Metric Invariance Test

To test whether respondents in different countries understood the survey questions similarly, author conducted a metric invariance test. Metric invariance tests examine whether the strength of the relationship between factors and items are the same across different groups of interest. The factor loadings are constrained to be equal across groups to evaluate model fit of the metric invariance using goodness-of-fit index criteria. Comparing this model in relation to the previous configural model, author determined if factor loadings are equal across groups. Several scholars (Byrne et al, 1989; Steenkamp & Baumgartner, 1998) have suggested that two equal factor loadings per each construct is enough to allow a comparison of effects, which is referred to as partial metric equivalence. Also, metric invariance is sufficient to compare constructs' unstandardized correlates with other variables of interest meaningfully across groups, such as covariances and unstandardized regression coefficients. Metric invariance is sufficient for the regression parameters to be comparable in a meaningful way.

Scalar Invariance Test

The scalar invariance test is employed to test equal indicator intercepts across groups (Meredith, 1993). If the scalar invariance is established, the means of the observed items are derived from the consequences of differences in the means of the constructs rather than from the differences in factor loadings or indicator intercepts (Steenkamp & Baumgartner, 1998). In this test, the factor loadings and factor intercepts are constrained to be equal across different groups. Thus, the findings of scalar invariance allow us to make meaningful mean comparisons of constructs. Comparing this scalar invariance model to the previous metric invariance models, author evaluates the model fit of the scalar invariance across countries. Several scholars also have suggested that two equal intercepts per construct are sufficient to compare means meaningfully (Byrne et al, 1989; Steenkamp & Baumgartner, 1998).

RESULTS

Single – Country Analyses

The proposed five factor model was specified and estimated repeatedly in an exploratory fashion until author obtains the best fitting model for each country of interest. Using multiple criteria for model fit, the best fitting model in each country of interest was obtained. Through the largest modification index and justification, the initial proposed five factor model was not accepted. Thus, after several re-specification and re-estimation processes, author developed the four factors model, dropping attachment to parent factor from the five factors model. Confirmatory factor analyses were conducted to test the adequacy of the four-factor structure using the data from each individual country in the ISRD-2. The four-factor model provided an adequate representation of the data in all participating countries, except for Portugal. Given this baseline model prerequisite for any subsequent measurement invariance tests, author decided to conduct further analyses with these 24 countries without Portugal. Overall, in all 24 countries, the four-factor model provided a satisfactory model fit. Thus, author could continue to conduct the further MGCEFA tests. According to the criteria introduced earlier, the CFI, RMSEA, and SRMR of the four factors model are acceptable for all participating countries in the study. Although there were small modifications in some countries, the results revealed that the four-dimensional model fits the ISRD-2 data across all countries.

Measurement Invariance Results

Taking a bottom-up approach, author gradually increased the number of constraints and assessed whether the model fit was not satisfied within the standard criteria when assessing configural, metric, and scalar invariance. When the model fit was not acceptable, author stopped the processes. Multiple group confirmatory factor analyses were conducted to answer the research question: are the instruments in the study invariant across different participating countries in the ISRD-2 study? According to Cheung and Rensvold (2002), the change in CFI between models should not exceed 0.01 when additional constraints are added in the next model. In addition, the change of RMSEA should not be exceeded 0.015 as well as 0.03 for SRMR in order for more constraints models to be invariant across groups. The results of the invariance tests are reported in the Table 1.

APPENDIX

Table 1: Measurement Invariance of the Four Factor Model in MGCF A (24 Countries (Portugal is deleted), N=44,221)**Table 1, part 1**

Model	Specifications	χ^2	Df	CFI
0	Configural Invariance	16785.35	6096	0.963
1	Full Metric Invariance	21933.45	6579	0.946
2	Partial Metric Invariance ^a	12906.24	4471	0.965
3	Scalar Invariance	52622.71	4965	0.803

Table 1, part 2

Model	Specifications	RMSEA	TLI	Δ CFI	Δ RMSEA
0	Configural Invariance	0.031	0.956	0.017	0.005
1	Full Metric Invariance	0.036	0.941	0.002	0.001
2	Partial Metric Invariance ^a	0.032	0.961	0.160	0.042
3	Scalar Invariance	0.073	0.800		

Configural Invariance

Configural invariance of the four factor models in CFA for all participating countries generally met at an acceptable level of model fit suggested for evaluation as: RMSEA and SRMR values close to .06 and .08 indicate acceptable fit, CFI values close to .95 indicate acceptable fit, and a χ^2/df ratio of 5:1 or less indicates a good fit. The final sample for multigroup confirmatory factor analysis in the current dissertation is 44,221 adolescents living in the 24 countries listed in Table 1. Maximum likelihood estimation was used to examine the configural invariance model as well as the other invariance models. Based on the results for the configural invariance base model (0 in Table 1), author cannot reject the configural invariance across different countries in the study. The goodness-of-fit statistics indicated a good fit for configural invariance: $\chi^2 = 16785.35$ (6096), $p < .01$, χ^2/df ratio = 2.75; RMSEA = 0.031; CFI = 0.963; TLI = 0.956.

Metric Invariance

The second model reports the fit indices for the metric invariance tests (1), which require equality between the factor loadings of the indicators across the participating countries in the study. Based on the results from the 1 (Table 1), the goodness-of-fit statistics indicated that the factor loadings were not invariant across the countries: $\chi^2 = 21933.45$ (6579), $p < .01$, χ^2/df ratio = 3.33; RMSEA = 0.036; CFI = 0.946; TLI = 0.941. Given that the modification indices suggested that the constraints on the item S1 (the first item of school attachment) and N1 (the first item of neighborhood attachment) should be dropped to improve model fit and to reduce nonequivalence across the countries in the study. After author checked the poor model fit from Model 1 (CFI = 0.954; TLI = 0.948), author dropped S1 and N1 with LSC1 (the first item of low self-control) and DP3 (the third item of delinquent peers) in 2. The results from 2 (Table 1) report that partial metric invariance was established: $\chi^2/df = 2.88$; SRMR = 0.039; CFI = 0.965; TLI = 0.961; Δ CFI = 0.002; Δ RMSEA = 0.001; Δ SRMR = 0.004. These results demonstrate that the partial metric invariance model fits the data as good as the configural model. Given that the partial metric invariance is obtained in the current study, the scale interval for the latent factor is partially equal across different countries. Consequently, difference scores, such as regression coefficients or covariances, could be compared cross-nationally.

Scalar Invariance

Next, I tested scalar invariance model (3 in Table 1). The results from scalar invariance model were not supported. The goodness-of-fit statistics did not indicate a good fit for scalar invariance: $\chi^2 = 52622.71$ (4954), $p < .01$, χ^2/df ratio = 10.62; RMSEA = 0.073; CFI = 0.803; TLI = 0.800. In summary, the results from the measurement invariance models were acceptable on the configural invariance and the partial metric invariance after making some modifications to indices. For the current dissertation's purpose, to compare difference score tests cross-nationally, the partial metric invariance is required. Consequently, based on the results from 0 to 2, regression coefficients across different countries can be compared in the next analyses.

CONCLUSION

The current study shows that testing measurement invariance is an important step for comparative analysis. Examining measurement invariance of instruments is critically important before comparing means or regression coefficients across different countries, because there is no way to know whether the differences in means across countries are due to actual differences in the latent construct or due to

measurement error. Taking a multi-group confirmatory factor analytic approach, the current study assessed the measurement invariance of four ISRD-2-scales that reflect various aspects from social bonding, self-control, and lifestyle and routine activities theories: attachment to school, attachment to neighborhood, low self-control, and delinquent peers.

The results of the current study contribute to our understanding of the importance of measurement invariance testing. However, some methodological challenges within current comparative research remain still unanswered by the current study. Thus, future research should expand the geographical range of included countries, and – if feasible – collect longitudinal data.

The findings reported in this study hopefully will provide momentum for making measurement invariance testing a common practice among comparative crime researchers.

REFERENCES

- Bennett, R. R. (2004). Comparative criminology and criminal justice research: The state of our knowledge. *Justice Quarterly*, 21(1), 1-21. doi:10.1080/07418820400095721
- Botchkovar, E. V., Tittle, C. R., & Antonaccio, O. (2009). General strain theory: Additional evidence using cross-cultural data. *Criminology*, 47(1), 131-176. doi:10.1111/j.1745-9125.2009.00141.x
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456-466. doi:10.1037/0033-2909.105.3.456
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233-255. doi:10.1207/s15328007sem0902_5
- Dammer, H. R., Reichel, P., & He, N. (2005). Comparing Crime and Justice. in *Handbook of Transnational Crime and Justice*, 23-46. doi:10.4135/9781452281995.n3
- Durkheim, E. (1951). *Suicide: A study in sociology*. Glencoe, IL: Free Press.
- Farrall, S., & Ditton, J. (1999). Improving the measurement of attitudinal responses: An example from a crime survey. *International Journal of Social Research Methodology*, 2(1), 55-68. doi:10.1080/136455799295186
- Farrall, S. (2004). Can we believe our eyes? A response to Mike Hough. *International Journal of Social Research Methodology*, 7(2), 177-179. doi:10.1080/1364557042000194568
- Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.
- Grasmick, H. G., Tittle, C. R., Bursik, R. J., & Arneklev, B. J. (1993). Testing the Core Empirical Implications of Gottfredson and Hirschi's General Theory of Crime. *Journal of Research in Crime and Delinquency*, 30(1), 5-29. doi:10.1177/0022427893030001002
- Karstedt, S. (2001). Comparing Justice and Crime across Cultures. *The SAGE Handbook of Criminological Research Methods*, 373-390. doi:10.4135/9781446268285.n25
- Meneses, R. A., & Akers, R. L. (2011). A Comparison of Four General Theories of Crime and Deviance: Marijuana Use Among American and Bolivian University Students. *International Criminal Justice Review*, 21(4), 333-352. doi:10.1177/1057567711408302
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543. doi:10.1007/bf02294825
- Miller, H. V., Jennings, W. G., Alvarez-Rivera, L. L., & Lanza-Kaduce, L. (2009). Self-control, attachment, and deviance among Hispanic adolescents. *Journal of Criminal Justice*, 37(1), 77-84. doi:10.1016/j.jcrimjus.2008.12.003
- Ren, L., Zhao, J. S., He, N. P., Marshall, I. H., Zhang, H., Zhao, R., & Jin, C. (2015). Testing For Measurement Invariance of Attachment Across Chinese and American Adolescent Samples. *International Journal of Offender Therapy and Comparative Criminology*, 60(8), 964-991. doi:10.1177/0306624x14566602
- Rodriguez, J. A., Perez-Santiago, N., & Birbeck, C. (2015). Surveys as cultural artefacts: Applying the International Self-Report Delinquency Study to Latin American adolescents. *European Journal of Criminology*, 12(4), 420-436. doi:10.1177/1477370815581701
- Steenkamp, J., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *J CONSUM RES*, 25(1), 78-107. doi:10.1086/209528
- Vazsonyi, A. T., Pickering, L. E., Junger, M., & Hessing, D. (2001). An Empirical Test of a General Theory of Crime: A Four-Nation Comparative Study of Self-Control and the Prediction of Deviance. *Journal of Research in Crime and Delinquency*, 38(2), 91-131. doi:10.1177/0022427801038002001