

Rethinking Measuring Moral Foundations in Prisoners: Validity Concerns and Implications

Hyemin Han ^{1*} and Mariola Paruzel-Czachura ^{2,3*}

¹ Educational Psychology Program, University of Alabama

² University of Pennsylvania

³ University of Silesia in Katowice

Author Note

Hyemin Han  <https://orcid.org/0000-0001-7181-2565>

Mariola Paruzel-Czachura  <https://orcid.org/0000-0002-8716-9778>

We have no known conflict of interest to disclose.

The authors have no funding to disclose.

Correspondence concerning this article should be addressed to Hyemin Han,
University of Alabama, Box 872031, Tuscaloosa, AL 35487, United States.

Email: hyemin.han@ua.edu

* Authors contributed equally to this work.

Rethinking Measuring Moral Foundations in Prisoners: Validity Concerns and Implications

Abstract

Prisoners, so those who probably engaged in criminal activities, might possess different perceptions and notions of moral foundations than non-prisoners. Thus, assessing such foundations among the population without testing the validity of the measure may produce biased outcomes. To address the potential methodological issue, we examined the validity of the measurement model for moral foundations among prisoners and community members, i.e., non-prisoners. We conducted the measurement invariance test and measurement alignment to test whether the model was consistently valid across the groups. We also employed the differential item functioning test to examine whether item responses were not biased between the different populations. Results demonstrated significant measurement non-invariance and differential item functioning. However, measurement alignment could address the non-invariance issue. Between-group comparisons of moral foundations were consistent with findings from prior research after performing the alignment.

Keywords: Moral foundations; Prisoners; Validity; Measurement invariance; Measurement alignment

Rethinking Measuring Moral Foundations in Prisoners: Validity Concerns and Implications

Moral foundations theory suggests that people prioritize different ethical values, which can be categorized into five distinct foundations (Graham et al., 2018). The care foundation is centered on empathy and concern for the suffering of others. Fairness involves a keen awareness of justice, rights, and equality. Loyalty reflects the tendency to form strong group bonds and take pride in group membership. Authority is about favoring structured social hierarchies and showing respect or deference to those in higher positions. The purity foundation is associated with disgust toward behaviors perceived as morally wrong, reflecting a concern for the sanctity of values. Care and fairness are considered individualizing foundations, as they focus on protecting individual rights. Loyalty, authority, and purity are binding foundations because they emphasize the preservation of group cohesion, traditional order, and shared values (Graham et al., 2011).

There is a long line of research showing that preference for moral foundations is related to immoral behaviors in community samples (e.g., Böhm et al., 2018; Milesi et al., 2020; E. Silver & Abell, 2016; Vecina, 2014), and also among prisoners (Paruzel-Czachura et al., 2023). The most consistent finding is that lower individualizing moral foundations are associated with harming behaviors. Moreover, prisoners had higher binding moral foundations than a control sample from the community (Paruzel-Czachura et al., 2023).

Although we possess some knowledge of moral foundations and immoral behaviors among the population, we still have concerns. We are particularly interested in whether moral foundations are measured validly among prisoners, who might possess different moral perceptions and notions than non-prisoner populations. Although there has been previous

research examining moral foundations among prisoners (e.g., Paruzel-Czachura et al., 2023), the methodology employed in such research, i.e., the MFQ, has only been tested and validated among non-prisoners (Graham et al., 2011). Quantitative research on moral foundations among prisoners, our population of interest, might be methodologically questionable without testing the measure's validity among them, whose perceptions of morality might differ from the general population.

Because more than 10 million people are incarcerated worldwide (National Institute of Corrections, 2015; Statista, 2021), and committing a crime brings social, economic, and moral negative consequences to society (Jones et al., 2019), we need to understand more moral foundations of people who not only do immoral behaviors but ended up in prison for their behaviors. That is why we aimed to understand the moral foundations of prisoners, particularly how those are measured, more deeply.

The Current Study

We examined whether the Moral Foundations Questionnaire (MFQ) (Graham et al., 2011) consistently measures five moral foundations as latent constructs across prisoners and non-prisoners. We already know prisoners have higher binding and lower individualizing moral foundations than control samples from the general population (Paruzel-Czachura et al., 2023). However, the study did not examine if prisoners would differently endorse and respond to items associated with foundations. This can be problematic when researchers intend to measure moral foundations among prisoners and compare the results with those from non-prisoner populations for further analysis. Because the moral foundation measure was primarily developed and validated among non-

prisoners, it is necessary to examine whether the measure assesses foundations among prisoners in a biased manner.

To address the concern, using a published database after the authors' permission (Paruzel-Czachura et al., 2023), we performed measurement invariance and differential item functioning (DIF) tests to test whether the potential differences while administering the MFQ across the two groups would be significant at the test and item levels. Our study will contribute to future studies in moral psychology to examine moral functioning, particularly moral foundations, among prisoners by addressing the potential concern of the biased measure. If any significant non-invariance or bias exists, the researchers will get practical insights into how to address them for appropriate moral foundation assessment among prisoners from the findings and discussions in the paper.

Materials and Methods

Participants

We analyzed a dataset collected from 764 participants in Poland. Among them, 50% were prisoners, while the rest 50% were non-prisoners. Specifically, 382 (136 women) prisoners aged from 19 to 71 ($M = 38.26$ years, $SD = 10.85$) and 382 (136 women) community members aged from 19 to 70 ($M = 30.45$ years, $SD = 9.94$). Prisoners reported their crimes; coded as violent or non-violent: $n = 109$ ($n_{\text{Men}} = 78$, $n_{\text{Women}} = 31$) were convicted of violence, and $n = 224$ ($n_{\text{Men}} = 136$, $n_{\text{Women}} = 88$) were convicted of non-violent crimes. Community members had not been convicted of a crime and were not accused of any when taking the survey. All participants were Caucasian by race (self-description) and of Polish nationality.

Materials

Moral Foundations Questionnaire

We employed the MFQ, measuring scores on five moral foundations (Graham et al., 2011). It consists of 30 items that measure the five moral foundations using two subscales where participants report their relevance (1 = *not at all relevant*; 6 = *extremely relevant*) or agreement (1 = *strongly disagree*; 6 = *strongly agree*) with the items measuring care, fairness, loyalty, authority, and purity. Responses to items were averaged to give an overall score for each foundation. We used its Polish version (Jarmakowski-Kostrzanowski & Jarmakowska-Kostrzanowska, 2016).

Statistical Analysis

With the permission of the authors of the main study (Paruzel-Czachura et al., 2023), we made all data and source code files available to the public via the Open Science Framework at <https://osf.io/k9a83/>.

Measurement Invariance Test and Alignment

We performed measurement invariance to test whether the measurement model of the MFQ is consistent and thus measures the foundations consistently across two groups (Putnick & Bornstein, 2016). We conducted a confirmatory factor analysis (CFA) and examined fit indicators, i.e., RMSEA, SRMR, and CFI, for testing invariance. In this process, four levels of invariance, i.e., configural, metric, scalar, and residual, were examined (see Supplementary Materials for detailed criteria for different levels of invariance) (Rachev et al., 2021). We assumed that data should at least support scalar invariance to perform further multigroup analysis, such as mean comparison.

If scalar invariance was not achieved, we performed measurement alignment to enable between-group comparisons of MFQ scores (Han, 2024; Robitzsch, 2024). We conducted measurement alignment after performing CFA for each MFQ subscale since measurement alignment is currently only available to a one-factor model (see Supplementary Materials for methodological details). After performing alignment, we examined resultant alignment indicators, $R^2_{loadings}$ ($\geq 95\%$), $R^2_{intercepts}$ ($\geq 95\%$), and % of items demonstrating non-invariance after alignment ($< 25\%$), to evaluate whether the procedure successfully addressed non-invariance (Asparouhov & Muthén, 2014).

Item Response Theory-based Test

We performed the DIF test to examine whether participants in different groups demonstrated significantly different item response probability when a similar foundation factor score was assumed. We used the *lordif* R package (Choi et al., 2011) to implement DIF testing with the Monte Carlo simulation (see Supplementary Materials for methodological details). Due to the large sample size, we focused on the effect size indicator for uniform and nonuniform DIF, R^2 , instead of p -values from χ^2 tests. The Monte Carlo simulation empirically determined the threshold for significant R^2 ($\alpha = .01$, iterations = 1,000). When R^2 exceeded the empirically determined threshold, we concluded that the DIF (either uniform or nonuniform) was not negligible (Choi et al., 2011).

Results

Measurement Invariance Test and Alignment

When we conducted a measurement invariance test, we found that the most lenient invariance, i.e., configural invariance, was not supported, RMSEA = .055, SRMR = .079, CFI = .760. RMSEA and SRMR sufficed the criteria, but CFI was lower than .900. Thus, we

performed measurement alignment to address the non-invariance. As reported in Table 1, measurement alignment successfully addressed the non-invariance issue for all subscales. Table 2 reports the results from mean score comparisons using aligned factor scores. Consistent with Paruzel-Czachura et al. (2023), we found that individualizing foundations scores (care, fairness) were lower among prisoners while binding foundations (loyalty, authority, purity) were higher than among community members.

Table 1

Measurement Alignment Performance Indicators

Foundation	$R^2_{loadings}$	$R^2_{intercepts}$	% loadings	% intercepts
Care	98.46%	99.77%	.00%	.00%
Fairness	99.64%	99.68%	.00%	16.70%
Loyalty	92.16%	99.40%	.00%	16.70%
Authority	98.93%	99.86%	.00%	.00%
Purity	97.31%	99.63%	.00%	.00%

Note. % loadings and % intercepts indicate the proportion of items demonstrating significant non-invariance in factor loadings and intercepts, respectively.

Table 2

Between-group Comparisons in Aligned Group Factor Means

	Non-prisoners		Prisoners		t	df	p	Cohen's d
	M	SD	M	SD				
<hr/>								

Care	.00	1.00	-.73	1.38	-8.36	762.00	< .001	-.61
Fairness	.00	1.00	-.90	1.31	-10.69	762.00	< .001	-.77
Loyalty	.00	1.00	.33	1.17	4.18	762.00	< .001	.30
Authority	.00	1.00	.70	.70	11.25	762.00	< .001	.81
Purity	.00	1.00	.31	.61	5.12	762.00	< .001	.37

Item Response Theory-based Test

The DIF test demonstrated that items demonstrated significant differential functioning across the two groups in all five subscales. As shown in Table S1, we found at least one item demonstrating uniform or nonuniform DIF in all five subscales. In such cases, the effect size indicator, R^2 , exceeded the empirically determined threshold estimated via Monte Carlo simulation.

Discussion

In the current study, the measurement invariance test reported that data does not support scalar invariance, so between-group comparison with simple composite scores could not be justified while comparing moral foundations across prisoners and non-prisoners (Han, 2024; Putnick & Bornstein, 2016). We could successfully address the non-invariance issue with measurement alignment. Group mean comparisons in five subscales replicated the findings from Paruzel-Czachura et al. (2023), i.e., prisoners reported lower moral foundations and higher binding moral foundations than participants from the community. Moreover, we found items demonstrating significant DIF in all five subscales. The result suggests that the item score probability is substantially different across two

groups when a similar moral foundation score is assumed, so the item-level measurement bias is not negligible and should be carefully treated by researchers (Choi et al., 2011).

The results suggest that when comparing moral foundations across people with different incarceration statuses, the current MFQ might be biased so measure foundations inconsistently. According to previous research employing the invariance and DIF test, significant non-invariance between groups (e.g., cross-cultural non-invariance) emerges when different groups understand and endorse items differently (e.g., Choi et al., 2019). This might also be the case while examining moral foundations among prisoners since they may possess different notions of morals and values (J. R. Silver & Silver, 2021; Vecina, 2014).

This issue requires further psychometrical treatments before conducting research projects on such between-group comparisons. First, researchers should test measurement invariance to examine whether the prerequisite for between-group comparison, scalar invariance, is ensured (Han, 2024; Putnick & Bornstein, 2016). The DIF test may also evaluate the item-level bias (Choi et al., 2019). Second, if significant non-invariance and item-level bias are discovered, researchers should consider performing measurement alignment to address the issue (Han, 2024). As mentioned in the introduction, measurement alignment is a feasible method to address the non-invariance issue, enabling between-group comparison by estimating adjusted group means. Another benefit of alignment is that it can address the item-level bias associated with significant DIF (DeMars, 2020).

Ethical Compliance

Funding: The authors have no funding to disclose.

Compliance with Ethical Standards: Not applicable (secondary analysis of previously collected and published data).

Conflicts of Interest: The authors declare they have no conflict of interest.

References

- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495–508.
<https://doi.org/10.1080/10705511.2014.919210>
- Brezina, T., & Piquero, A. R. (2007). Moral beliefs, isolation from peers, and abstention from delinquency. *Deviant Behavior*, 28(5), 433–465.
<https://doi.org/10.1080/01639620701233324>
- Böhm, R., Thielmann, I., & Hilbig, B.E. (2018). The brighter the light, the deeper the shadow: Morality also fuels aggression, conflict, and violence. *The Behavioral and Brain Sciences*, 41, e98. <https://doi.org/10.1017/S0140525X18000031>
- Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39(8), 1–30. <https://doi.org/10.18637/jss.v039.i08>
- Choi, Y. J., Han, H., Dawson, K. J., Thoma, S. J., & Glenn, A. L. (2019). Measuring moral reasoning using moral dilemmas: evaluating reliability, validity, and differential item functioning of the behavioural defining issues test (bDIT). *European Journal of Developmental Psychology*, 16(5), 622–631.
<https://doi.org/10.1080/17405629.2019.1614907>

DeMars, C. E. (2020). Alignment as an alternative to anchor purification in DIF analyses.

Structural Equation Modeling: A Multidisciplinary Journal, 27(1), 56–72.

<https://doi.org/10.1080/10705511.2019.1617151>

Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitch, C., & Mooijman, M. (2018). Moral foundations theory: On the advantages of moral pluralism over moral monism. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 211–222). The Guilford Press.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2), 366–385.

<https://doi.org/10.1037%2Fa0021847>

Han, H. (2024). Using measurement alignment in research on adolescence involving multiple groups: A brief tutorial with R. *Journal of Research on Adolescence*, 34(1), 235–242. <https://doi.org/10.1111/jora.12891>

Milesi, P., Süßenbach, P., Bohner, G., & Megías, J.L. (2020). The interplay of modern myths about sexual aggression and moral foundations in the blaming of rape victims. *European Journal of Social Psychology*, 50, 111–123. <https://doi.org/10.1002/ejsp.2622>

Paruzel-Czachura, M., Blukacz, M., Vecina, M. L., & Jonason, P. K. (2023). Moral foundations and criminality: Comparing community members to prisoners and violent/non-violent offenders. *Psychology, Crime & Law*.

<https://doi.org/10.1080/1068316X.2023.2246624>

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research.

Developmental Review, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>

Rachev, N. R., Han, H., Lacko, D., Gelpí, R., Yamada, Y., & Lieberoth, A. (2021). Replicating the Disease framing problem during the 2020 COVID-19 pandemic: A study of stress, worry, trust, and choice under risk. *PLoS One*, *16*(9), e0257151.

<https://doi.org/10.1371/journal.pone.0257151>

Robitzsch, A. (2024). Examining Differences of Invariance Alignment in the Mplus Software and the R Package Sirt. *Mathematics*, *12*(5), 770.

<https://doi.org/10.3390/math12050770>

Silver, E., & Abell, L. (2016). Beyond harm and fairness: A study of deviance and morality. *Deviant Behavior*, *37*, 496–508. <https://doi.org/10.1080/01639625.2015.1060746>

Silver, J.R., & Silver, E. (2021). The nature and role of morality in offending: A moral foundations approach. *Journal of Research in Crime and Delinquency*, *58*, 343–380.

<https://doi.org/10.1177/0022427820960201>

Vecina, M.L. (2014). The Five Moral Foundations Sacredness Scale in men in court-mandated treatment for violently abusing their partners. *Personality and Individual Differences*, *64*, 46–51. <https://doi.org/10.1016/j.paid.2014.02.021>