

Adjusting Group Intercept and Slope Bias in Predictive Equations

Bruce W. Austin^a, Brian F. French^a

[a] Learning and Performance Research Center, Department of Kinesiology and Educational Psychology, Washington State University, Pullman, WA, USA.

Methodology, 2020, Vol. 16(3), 241–257, <https://doi.org/10.5964/meth.4001>

Received: 2019-06-04 • **Accepted:** 2020-06-04 • **Published (VoR):** 2020-09-30

Corresponding Author: Bruce W. Austin, Cleveland Hall 364, PO Box 642136, Pullman, WA 99164-2136, USA. E-mail: bwaustin@wsu.edu

Abstract

Methods to assess measurement invariance in constructs have received much attention, as invariance is critical for accurate group comparisons. Less attention has been given to the identification and correction of the sources of non-invariance in predictive equations. This work developed correction factors for structural intercept and slope bias in common regression equations to address calls in the literature to revive test bias research. We demonstrated the correction factors in regression analyses within the context of a large international dataset containing 68 countries and regions (groups). A mathematics achievement score was predicted by a math self-efficacy score, which exhibited a lack of invariance across groups. The proposed correction factors significantly corrected structural intercept and slope bias across groups. The impact of the correction factors was greatest for groups with the largest amount of bias. Implications for both practice and methodological extensions are discussed.

Keywords

invariance, noninvariance, predictive bias, test bias, intercept bias, slope bias, multigroup

The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) provide guidelines for all facets of testing from the statistical properties of a test to how results are used to assist public policy or employment decisions. The *Standards* directly address the ethical issues involved in fair and unbiased testing. Many guidelines that concern the misuse of tests were developed, in part, as a result of Title VII of the Civil Rights Act of 1964 in the United States, which provided government regulations to ensure that employment decisions were not based on a person's ethnicity, race, sex, or religion.



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License, CC BY 4.0](#), which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.

Measurement invariance (MI) is an essential component for constructing a validity argument and a precursor to score use. Validity evidence related to associations with external variables is especially important to support inferences across groups. MI effects are documented for mean differences and group comparisons (Ferne & Rupp, 2007). However, less attention has been given to how a lack of MI influences predictive validity (Millsap, 2011). A lack of MI between groups can cause bias in the intercept and slope of a common regression equation (Aguinis, Culpepper, & Pierce, 2010). This differential prediction or test bias (Crocker & Algina, 1986) can result in over- or under-prediction of an outcome for one group relative to another. This work focuses on these predictive relationships in the absence of MI, set in the context of cross-cultural comparisons, as this context represents fertile ground for understanding how bias operates across groups. We propose and demonstrate group-level correction factors for regression (i.e., structural) intercept and slope bias in common regression equations, where the bias likely resulted from a lack of factor invariance (FI), a form of MI, in an international dataset.

Item and factor structure differences may be present as a result of cultural differences (Church et al., 2011). The presence of such construct irrelevant variance threatens score validity, rendering cross-cultural comparisons problematic, (Hancock, 1997). At the item level, for example, Church et al. documented differential item functioning (DIF) across cultures in up to 50% of the items in the revised NEO (Neuroticism, Extraversion, Openness) personality inventory that influenced facet scores (NEO; McCrae et al. 2005). This issue may be exacerbated by the fact that many widely used personality and intelligence measures were developed in the United States (US). The interpretation and use of these scores from an instrument based in English and the US culture can have negative consequences on individuals if variability resulting from translation and culture is not controlled.

Factor invariance can be examined at the factor structure level through multigroup confirmatory factor analysis (MGCFA; Millsap, 2011). Millsap provides a complete description of the levels of FI including configurational, metric and scalar. The presence of FI implies that the latent variables are measured in the same manner for subgroups examined, and that scores on the observed manifestation of the latent variable are the same for members of different groups with the same level of the measured trait.

Several options exist when analyzing data that lack FI. Researchers can ignore the problem or estimate non-equivalence models (Kuha & Moustaki, 2015) and argue for valid group comparisons under partial invariance (Steenkamp & Baumgartner, 1998). Ignoring a lack of FI can produce predictive models where one group is unfairly favored and group comparisons are problematic from a theoretical and conceptual perspective. If a researcher estimates a non-equivalence model (e.g., certain items differ across groups), similar theoretical and conceptual problems of construct equivalence are encountered. Furthermore, with a large number of groups, partial invariance models become less

justifiable given problems finding an invariant referent for the latent variable (French & Finch, 2008). Neither ignoring a lack of FI, nor estimating non-equivalent models, can be justified when making decisions about people. Kuha and Moustaki (2015) concluded that the sensitivity of group comparisons under non-equivalent measurement can be severe, leading to biased conclusions.

Kim et al. (2017) examined five methods for testing groups for measurement invariance. Kim found that the alignment method was adequate for establishing approximate invariance but is not recommended when “many measurement parameters are substantially noninvariant” (p. 539). Lomazzi (2018) examined the alignment procedure for polytomous items and achieved an “acceptable degree of noninvariance” for 35 out of 59 country groups, a 59% success rate. This approach produced a limited model that was not applicable in 40% of the groups in the examined dataset. Another option for prediction is needed that can address noninvariance in all groups in a dataset and that can minimize the need to make judgements about degrees and combinations of measurement noninvariance across items.

Predictive Invariance

Assumptions about regression or structural slope and intercept bias in predictive equations has been questioned. Aguinis, Culpepper, & Pierce (2010) found that structural intercept bias is likely over-estimated while bias from structural slope differences goes undetected using well established procedures. Structural slope bias was especially pernicious due to great difficulty in detecting it. Power to detect slope-based bias was found to be less than 10% in samples as large as 400 under conditions commonly encountered in social science research (Aguinis et al., 2010; Pokropek et al. 2019).

Latent variables and factor scores are commonly used in predictive equations. Unfortunately, the predictive invariance of such equations can be compromised by the use of assessments that lack invariance across groups (Millsap, 2011). For example, if predictor measurement intercepts in a factor model are consistently greater for one group compared to another, then the resulting predictive bias may lead to one group being systematically favored in the outcome. Furthermore, group mean differences can confound the effects of measurement noninvariance on predictive equations. A lack of invariance in the predictor measurement intercepts can be cancelled out by group mean differences, producing what appears to be an invariant predictive equation. In such a situation, one group will be systematically favored over the other even though the predictive equation exhibits invariant structural slopes and intercepts. In addition, factors such as reliability, group sample size proportions, correlation between criterion and group, differences in predictor variances by group, invariance of the criterion variable can all work together to produce predictive bias in structural slopes and intercepts (Aguinis et al., 2010).

Detection of invariance in a model has received much attention (Rutkowski & Svetina, 2014). Threshold criteria (Chen, 2007; Cheung & Rensvold, 2002) have been

cited extensively, yet are dependent on limited research in how various conditions, such as sample size and uniformity of noninvariance, impact the validity of these thresholds. This uncertainty leads to the problematic factor of subjective judgments about whether a study's characteristics are similar enough to models examined in simulation on which threshold criteria were established. Furthermore, these criteria only offer a dichotomous decision about whether invariance is supported and little information on the magnitude and impact of such differences.

Numerous combinations of noninvariance can occur among measurement intercepts and slopes for both criterion and predictor variables. These combinations, along with differences in group population means, can mask predictive bias (Millsap, 2011). We follow Millsap's recommendation to continue the study of this issue. We focus attention on FI in the predictor variable and connecting a lack of FI to predictive equations. Our aim is to give researchers tools to meet their responsibility for accurate comparisons.

Purpose of the Study

We developed two correction factors (CF), one for intercept and one for slope in a common regression equation, to adjust bias between groups that originated from a lack of FI at the item level. These CFs were designed to (a) adjust bias *at the group level* and not shift bias from one group to another, (b) be used within a latent variable context, and (c) be easy to implement with many identifiable groups. We demonstrated how use of the CFs can increase the accuracy of latent variable predictors. We note this is not a permanent fix but a method to allow data to be used for fair analyses. The underlying cause of bias must be addressed for future data collection. However, it may serve to improve accuracy with existing data analysis.

The primary benefits of this method are (a) a more objective approach to adjusting predictive bias that precludes the need for subjective decisions, (b) not relying on the use of fit statistics and thresholds (i.e. Chen's, 2007 fit criteria) for models and data that likely do not meet the narrow simulated characteristics of the models from which the fit criteria were obtained, (c) a quantification of predictive bias in terms of the specific values of the slope and intercept CFs for each group, (d) the use of a single predictive model (with CFs as additional inputs), and (e) the testing of statistically significant differences between the CFs of different groups.

We hypothesized that the CFs would allow for greater predictive accuracy compared to prediction without the corrections. We also hypothesized that predictive bias would be controlled for to a greater extent with groups, where FI was not present in the predictor variable compared to their counterparts where FI was present.

Method

Our overall procedure was to first examine the FI of a construct at the configural, metric, and scalar levels. Second, we constructed our CFs. Third, we tested the CFs in a regression equation. We used Mplus (Version 7.4) for all analysis (Muthén & Muthén, 2012).

Sample

We used the 2012 Programme for International Student Assessment (PISA) student dataset ($N = 485,490$; The Organisation for Economic Co-operation and Development [OECD], 2014) that is completely anonymous and publicly available. The dataset contains responses to a questionnaire and mathematics test from 15-year-old students in 68 countries and regions.

Measures

We selected a set of variables to construct a latent variable regression model to illustrate the CF method. The PISA math achievement score (PVMATH, $\alpha = 0.91$) was the dependent variable and mathematics self-efficacy, an 8-item scale ($\alpha = 0.84$), was the independent variable. The PISA math achievement score was selected as the criterion to minimize any possible confounding effects that might result from the use of a psychological construct that also suffered from a lack of FI. Furthermore, regardless of whether a single factor or two factors underlie the criterion and predictor, there is no difference in the relationship between predictive slope invariance and predictor FI (Millsap, 1997).

Reference Group Selection

Given the negative effect of different reference and target group sample size proportions on predictive intercept bias (Aguinis et al., 2010; Chen, 2007), as well as the individual vagaries of any single country in the dataset, we did not select one country as the reference group. Instead, we randomly selected a set of 4,530 records from across the dataset to form a random reference group that approximated the average sample size of the 68 countries and regions and would thus minimize bias due to different group proportions. We modeled this process after the use of calibration samples by the OECD where cases were randomly selected across all countries for obtaining international item parameters (OECD, 2014).

Multigroup Confirmatory Factor Analysis

We examined the mathematics self-efficacy items for metric and scalar invariance between each country and a 1-factor baseline model established by the reference group. Every country was compared with the reference group using progressively constrained

models (Bollen, 1989). Chen's (2007) fit index criteria were used to determine a lack of invariance which make use of concurrent changes in CFI (comparative fit index), RMSEA (root mean square error of approximation), and SRMR (standardized root mean square residual).

Structural Intercept Correction Factor

The structural intercept correction factor was created by first calculating an estimate of the predictive intercept bias for each group in our dataset using an equation from Aguinis et al. (2010, p. 653). We refer the reader to the Appendix of the Aguinis et al. article for a proof of the formula. Aguinis' equation is an absolute measure of bias which lacks information about whether the common regression intercept over- or under-estimates the group intercept. Therefore, we added a directional component taken from the sign of the correlation of the criterion and group. [Equation 1](#) shows the resulting structural intercept correction factor that produces a unique value for each group in the dataset.

$$Adj\Delta I_G = SNr_{yG} \left[\frac{(r_{yG} - r_{xyG}\sqrt{p_{rG}(1-p_{rG})}\Delta\mu_{rG})^2}{1 - p_{rG}(1-p_{rG})(\Delta\mu_{rG})^2} \right] \beta_0 \quad (1)$$

where $Adj\Delta I_G$ denotes the correction factor for group G, SNr_{yG} is the sign of the correlation of the criterion and group G, r_{yG} is the correlation between the criterion and group G, r_{xyG} is the correlation between the predictor and criterion in group G, p_{rG} is the proportion of the group G sample size to that of the reference group, $\Delta\mu_{rG}$ is the difference in predictor means between group G and the reference group and β_0 is the common regression line intercept.

The correction factor can then be used in a predictive equation such as a common regression line for all groups in the dataset. Hence, the new common regression line appears as in [Equation 2](#), where Y_{adj} designates the adjusted criterion.

$$Y_{adj} = Adj\Delta I_G\beta_0 + \beta_0 + \beta_1X + \epsilon \quad (2)$$

The structural slope correction factor was calculated as a function of the ratio of the target and reference group predictor factor variances (Millsap, 1997, p 254). The ratio of the predictive slopes of two groups is equivalent to the ratio of their communalities. We refer the reader to Millsap (1997) for a complete exposition of this formula. We first calculated the variance of the factor for the reference group alone, and then calculated the factor variance for each of the 68 groups using the factor scores for the predictor calculated using the entire dataset. The slope adjustment factor was then calculated as the ratio of the difference in target and reference group variances in terms of the reference group as in [Equation 3](#). As with the intercept adjustment component, the sign

of the correlation of the criterion and group provided a needed directional component to the equation.

$$Adj\Delta S_G = SNr_{yG} \left[\frac{\sigma_G^2 - \sigma_{reference}^2}{\sigma_{reference}^2} \right] \quad (3)$$

where SNr_{yG} is the sign of the correlation of the criterion and group G, σ_G^2 is the variance for group G, and $\sigma_{reference}^2$ is the variance for the reference group.

Equation 2 is further modified by adding the slope correction factor (Equation 3) to produce Equation 4, the adjusted regression line, allowing for simultaneous correction of slope and intercept bias.

$$Y_{adj} = Adj\Delta I_G \beta_0 + \beta_0 + (\beta_1 - Adj\Delta S_G \beta_1)X + \varepsilon \quad (4)$$

Assessing the Correction Factors

Following recommendations by Aguinis et al. (2010) and standards from AERA, APA, and NCME (2014), we assessed the correction factors by comparing their predictions, using Equation 4, with results from regression lines calculated independently for each individual country. We first constructed a common latent variable regression model by regressing math achievement on the math self-efficacy construct. We then estimated the same regression model individually for each country. Predicted values from the common and individual models could then be compared and bias assessed. Predictive bias could then be calculated as the absolute difference between the two sets of predicted values on a country by country basis.

The common regression model was then modified using the correction factors (Equation 4) which adjusted the predicted values on a group-by-group basis. The predicted values from the adjusted regression model could then be compared with the values from the individual regression models and bias re-assessed. In theory, the predicted values from the *adjusted* common regression model should match, or closely approximate, the values obtained from the individual regression models. A close approximation would indicate an elimination or reduction in predictive bias for all groups without simply transferring it from one group to another in line with our goal to eliminate bias *at the group level*.

Results

MGCFA Results

Metric FI held for the mathematics self-efficacy items across all countries, providing no guidance for where predictive slope bias would eventually be found between our common regression and the individual country regressions. A lack of scalar FI appeared

in 32 countries. We would expect, therefore, that countries exhibiting a lack of scalar invariance would also exhibit greater predictive intercept bias than the countries where scalar FI held. MGCFCA model results are presented in [Supplementary Materials, Excel 1](#).

Correction Factor Results

Table S1 in the [Supplementary Materials](#) displays the slope and intercept bias present in the common regression line (Columns 2 and 3) along with the average bias obtained by estimating a set of predicted values in the common regression line (Column 4) and the adjusted regression line (Column 5). Column 1 of the table lists the countries and regions. Column 2, Slope Bias, is calculated as the difference between the individual regression and common regression slopes (unstandardized) as a percentage of the individual regression slope. For example, the slope of the common regression line ($b = -108.252$) is 20.79% smaller in magnitude than the slope of the individual regression line for Australia ($b = -136.657$). Column 3 in Table S1 (see [Supplementary Materials](#)) indicates the intercept bias present in the common regression line for each country. Intercept bias is calculated as the difference between the common regression predicted value, estimated using the math self-efficacy mean for each country, and the individual country regression intercept, as a percentage of the individual regression intercept. For example, the common regression line underestimates the country mean for Australia by 4.79%. Column 4 presents the average predicted value bias in the unadjusted common regression line while the last column in Table S1 (see [Supplementary Materials](#)) shows the average predicted value bias of the adjusted common regression line with the intercept and slope correction factors. The average predicted value bias for Columns 4 and 5 are calculated as the absolute differences in predicted values averaged across five points along the math self-efficacy scale. These five points are the M , $-2SD$, $-1SD$, $+1SD$, and $+2SD$. This selected range of predicted values provides a way to compare predicted values between the common and adjusted regression lines. Again, using Australia, there was an average bias in predicted values produced by the *unadjusted* common regression line of 4.89% (Column 4) compared with the individual regression line for Australia. The 4.89% bias was reduced by 46% to 2.65% (Column 5) using the correction factors in the *adjusted* common regression line.

The average magnitude of the predictive (structural) slope bias (Table S1, Column 2 in [Supplementary Materials](#)) was 68%, with a range in magnitude from as little as 0.28% (Korea) to 2325.0% (Albania). Furthermore, in the 57 regions where bias was reduced, the intercept and slope correction factors reduced bias in the predicted values from the *unadjusted* common regression line by 57.4%, on average, with a range of 7.4% to as high as 97.5%. Interestingly, larger reductions in bias were obtained where bias was large between the individual and common regression lines. Bias was reduced by 66.6%, for example, in all countries where the original unadjusted bias was greater than 7%. In the 11 regions where the correction factors did not reduce bias, average bias was only 3.27%.

The average magnitude of *unadjusted* predictive bias (average of Table S1, Column 4 in [Supplementary Materials](#)) across all countries was 7.8%, but 9% across the countries lacking scalar invariance according to [Chen's \(2007\)](#) criteria. After applying the structural slope and intercept correction factors, the average magnitude of predictive bias (average of Table S1, Column 5 in [Supplementary Materials](#)) was reduced to 3.2%.

All calculations for the results in Table S1 are provided in [Supplementary Materials](#), [Excel 2](#).

Structural Intercept Correction

Disaggregating results by each correction factor, the intercept correction factor reduced structural intercept bias from 1% to nearly 100%. Figure S1 in [Supplementary Materials](#) illustrates the relationship between the amount of structural intercept bias present in the common regression line for the 68 countries (solid line corresponding to Intercept Bias column in Table S1 in [Supplementary Materials](#)) and the impact of the structural intercept correction factor to reduce that bias (dashed line). The larger the original bias, the larger the effect of the structural intercept correction factor. Figure S1 (see [Supplementary Materials](#)) illustrates that Indonesia, for example, showed structural intercept bias of 25.02%, one of the highest, while the structural intercept correction factor reduced that bias by 96.5%. In countries with small structural intercept bias, such as Slovenia (3.08%), the reduction tended to be much smaller. For example, Slovenia's bias was only reduced by 7%. A regression analysis of the relationship between the percent reduction and the original amount of common intercept bias was significant ($p < .05$, $R^2 = 78\%$). For every percent increase in structural intercept bias, the correction factor reduced bias by 4.4%. The countries where the structural intercept correction performed poorly (e.g. Liechtenstein) displayed individual regression line slopes that deviated from the common regression line slope much more than countries where the intercept correction factor did perform well.

Structural Slope Correction

Focusing on the contribution of the structural slope correction factor, the structural slope correction reduced predictive bias in 70% of the countries beyond the reduction obtained with the structural intercept correction alone. Figure S2 in [Supplementary Materials](#) compares the original unadjusted predictive bias (solid line) with the predictive bias after incorporating the structural intercept adjustment factor (dashed line) and the bias after adding the structural slope adjustment factor *with* the structural intercept adjustment (dotted line). The figure illustrates how the slope adjustment factor, in conjunction with the intercept correction, consistently reduced bias across countries and was effective when the original unadjusted bias was large (right hand side of the figure). There were, however, situations where the slope adjustment factor did not reduce predictive bias.

As with our examination of the intercept correction factor, we examined the slope correction with cases where the *original predictive bias* was greater than 7%. These countries exhibited large structural slope bias, 135% on average, as opposed to 18% slope bias for the countries with less than 7% predictive bias. Table 1 displays the sample break down in bias reduction by structural intercept and slope correction factors for the countries where the original predictive bias was *Greater Than 7%*, *Less Than 7%*, and the whole sample. Overall, the structural slope and intercept bias correction factors reduced average predictive bias from 7.79% to 3.22%. For the *Less Than 7%* group, average predictive bias was only 4.31% and the intercept correction factor reduced most of that predictive bias (from 4.31% down to 3.61%) with the addition of the slope correction factor only reducing predictive bias down to 3.06%. For the *Greater Than 7%* group, however, the intercept correction factor reduced predictive bias by more than half (12.3% to 4.97%) with the addition of the slope correction factor further reducing predictive bias to 3.44%.

Table 1*Breakdown of Predictive Bias Reduction by Average Amount of Original Predictive Bias*

Breakdown of averages	Original unadjusted predictive bias (%)	Predictive bias after intercept adjustment (%)	Predictive bias after intercept and slope adjustment (%)
Average predictive bias across all 68 countries	7.79	4.19	3.22
Average across countries with original bias LESS THAN 7%	4.31	3.61	3.06
Average across countries with original bias GREATER THAN 7%	12.48	4.97	3.44

We note that the countries and regions where a lack of FI was identified using Chen's (2007) criteria (starred in Table S1, see *Supplementary Materials*) had a higher percent of predictive bias on average (i.e., 9.8%) in comparison to their counterparts where FI was present (i.e., 6.0%). In addition, the intercept and slope correction factors for the countries and regions that lacked FI reduced bias to an average of 3.24%, which is comparable to the remaining bias in the groups where FI was present (i.e., 3.2%). Moreover, the average reduction in predictive bias was larger than for the lack of FI groups (66.9%) than for the FI groups (46.8%). This supports that the structural intercept correction factor can adjust for issues in the measurement of the predictor but perhaps not all sources of bias. A regression model of the relationship between the percent reduction and the original amount of predictive bias, using both the slope and intercept correction factors,

was significant ($p < .05$, $R^2 = 50\%$). For every percent increase in predictive bias, the correction factors reduced bias by 36%.

Country Examples

The results for Brazil, Macao, and Liechtenstein are outlined below to illustrate examples of how the correction factors work at the group level. First, the common regression equation calculated across all countries is shown in [Equation 5](#).

$$Y = 469.652 + (-108.252)X \quad (5)$$

The intercept and slope adjustment values for Brazil are -0.1426 and 0.4569, respectively. [Equation 6](#) illustrates the inclusion of Brazil's adjustment values in [Equation 4](#). Thus, predicted values for all members of group Brazil will be calculated using [Equation 6](#) where b_0 and b_1 are the coefficients of the common regression line calculated across all groups. All calculations are performed using the unrounded values (see [Supplementary Materials, Excel 2](#)).

$$Y_{adj} = -0.1426b_0 + b_0 + (b_1 - (0.4569)b_1)X_{Brazil} \quad (6)$$

Inclusion of the coefficients from the common regression line into [Equation 6](#) produces [Equation 7](#) for group members of Brazil.

$$Y_{adj} = 402.675 - 58.79X_{Brazil} \quad (7)$$

Similar calculations for group members of Macao and Liechtenstein produce [Equation 8](#) and [Equation 9](#), respectively.

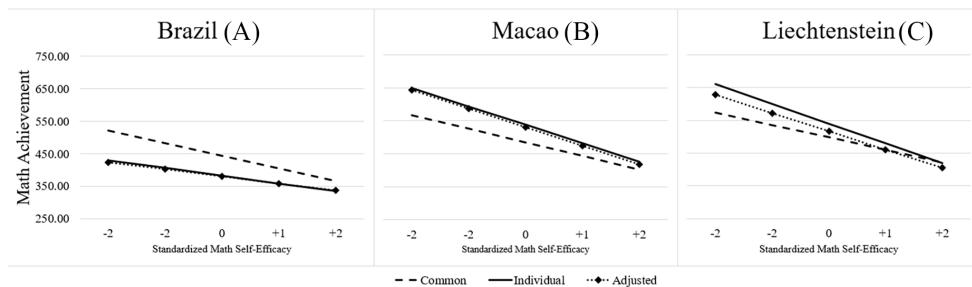
$$Y_{adj} = 507.391 + (-150.146)X_{Macao} \quad (8)$$

$$Y_{adj} = 474.158 + (-159.507)X_{Liechtenstein} \quad (9)$$

The panels shown in [Figure 1](#) illustrate the reduction in predictive bias for Brazil, Macao, and Liechtenstein using both the slope and intercept correction factors in the common regression line. The correction factors produced predicted values *as if the common regression line was effectively moved to overlap the individual country regression lines* on a country-by-country basis. Please note that the adjusted regression line for one country is not the same as the adjusted regression line for another because the slope and intercept correction factors are unique for each country as illustrated in the country examples above. The figures simply illustrate the *effective change* in the common regression line via the correction factors.

Figure 1

Comparison of Common, Individual, and Adjusted Regression Lines Using Both the Intercept Adjustment Factor and the Slope Adjustment Factor



Note. The adjusted line represents how predicted values are produced from the common regression line as if it was effectively adjusted to match the individual country regression line for each group.

Figure 1, Panels A and B reflect corrections to the common regression line in terms of predicted values when using both correction factors. Individual country predicted values are effectively produced as if the common regression line (dashed line) was adjusted (dotted diamond line) to almost perfectly overlap the individual regression line (solid line). Panel C also illustrates a reduction in predictive bias where the intercept adjustment alone failed to improve predicted values. While not as good a correction as seen in Panels A and B, the slope of the adjusted regression line for Panel C more accurately represents the true relationship between the predictor and criterion in Liechtenstein.

Discussion

This study established and demonstrated two correction factors for predictive bias found in common regression lines estimated using group data from a large-scale international dataset with a lack of FI in the predictor variable. The aim was to provide researchers with a third option to account for bias beyond ignoring it or using partial invariance models and that minimizes the need to make judgements about the presence and degree of measurement noninvariance. The method we propose, while applicable in its current form, is not necessarily intended to be a definitive answer to this complex problem, but a new approach, open to further development, that synthesizes a wide variety of information (e.g. Aguinis et al., 2010; Millsap, 1997) not previously utilized in such a manner.

The correction factors produced large and consistent adjustments to *predicted values* on a group-by-group basis, obtained using a common regression line, estimated across all countries in the dataset. The corrections approached 100% of the estimated bias in

predicted values, especially in cases where bias was large (i.e. 7% to 22%, in 31 of 68 regions). This is a considerable benefit, especially when common predictive equations are used to make decisions about people and resources in contexts such as education or employment. Moreover, the reduction in predictive bias was similar for countries identified with a lack of FI according to [Chen's \(2007\)](#) criteria (32 regions) compared with those where FI was present (36 regions). In countries with a lack of FI, bias was reduced from an average of 9.7% to 3.2%. For countries with FI present, bias was reduced from 6.0% down to 3.2%. This supports that the correction can account for, in part, issues in the measurement of the predictor variables. Given that some bias remained, other sources (e.g., unreliability, restriction of range) still need to be studied and accounted for in such correction factors.

The performance of the correction factors also suggests that over-correction for bias is fairly minimal and limited to cases where predictive bias is less than 5% (Figure S2 in [Supplementary Materials](#)). The results show that the effect of the correction factors seems to diminish as predictive bias gets smaller. Future research should address over-correction by the slope correction factor in the small number of cases where it occurs. Cases where over-correction occurred will provide useful information for future research such as helping to establish parameters to guide simulation work and address the main limitation of this study: a case study of a single construct from one dataset. Such evidence is useful given that slope bias is likely present, yet difficult to detect due to low power, even with large sample sizes ([Aguinis et al., 2010](#)).

The use of the correction factors addresses standards (e.g., [AERA, APA, & NCME, 2014](#)), for ensuring fairness and equity in decisions about individuals from different groups. Whether such groups consist of different races/ethnicities or different nationalities and cultures, the correction factors can be applied where groups are identified, bias is shown to exist, and revision of assessments is not possible.

Both correction factors are easy to create and do not require advanced statistical knowledge to implement. Calculation of factor scores from latent variables, reliability, sample proportions, and factor variances are all elements that are easily obtained from software packages capable of latent variable estimation. The correction factors should be helpful in large cross-cultural or international datasets where the number of groups is large enough that the use of group indicator variables, or partial invariance techniques, is impractical. We further recommend that researchers use MGCFA and established threshold-based invariance measures (e.g. [Chen, 2007](#)) to document sources of predictive bias while being mindful that structural slope bias is difficult to detect ([Aguinis et al., 2010](#)). Individual group regression lines should be estimated to assess the performance of the correction factors and assure standards of fairness are met. Basic steps to follow include (a) identify groups with the potential for predictive bias (e.g. countries), (b) use the groups to perform an MGCFA, (c) estimate common and individual regression lines, (d) assess levels of predictive bias, (e) create intercept and/or slope correction factors, (f)

apply the correction factors to the common regression line, and (g) assess results using predicted values from the regression outcome.

Limitations and Future Directions

The implications of the statistical bias (approximately 4%) in Aguinis' et al. (2010) bias formula, used for our intercept correction factor, is not clear. Given that the bias in the formula is driven mainly by unreliability of test scores, perhaps that can be integrated into the correction factors.

Given our data source, we had to assume that the criterion was free from measurement problems. While a reasonable assumption given the variable, we could not verify it. The effects of a lack of FI on both the predictor and the criterion while using correction factors deserves attention.

To address such limitations, a series of simulation studies is recommended. First, a diverse set of conditions should manipulate elements that are used in our correction factors in addition to other influences (e.g., restriction of range) to understand how the correction factors function under a range of conditions where bias is known and varied, especially intercept and slope bias. Second, such work can help understand if the adjustments over- or under-correct for bias and if reference and focal groups are being misrepresented through such corrections. Third, exploration of different reference group selection criteria, such as comparing our random selection method with the use of an established group (e.g. country or language group) as the reference is needed. Fourth and finally, comparisons of this method with other methods such as the Alignment Method, and recent work using regularized nonlinear multigroup factor analysis for invariance (e.g., Bauer et al., 2020) would highlight strengths and weaknesses of such adjustments for different situations.

Thus, given the legal and practical implications of adjusting data and the unknown aspects of the adjustments, we cannot recommended it for use without further evaluation. However, we do encourage continued research in this area to build a stronger analytical and empirical connection between measurement issues and predictive bias.

Conclusion

A lack of measurement invariance, specifically FI, in predictor variables can have a cascading effect on predictive equations resulting in differential prediction or test bias. This can have meaningful implications for individuals and groups, with some being unfairly favored over others. Our proposed correction factors have the potential to extend well-beyond educationally-related variables in cross-cultural settings. These methods may be useful for any group comparisons for a variety of inferences that need support.

Funding: The authors have no funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Acknowledgments: The authors have no support to report.

Supplementary Materials

For this article the following supplementary materials are available ([Austin & French, 2020](#)):

- Table S1: containing detailed country-level results.
- Figure S1: illustrating the corrections for structural intercept bias.
- Figure S2: illustrating the combined structural intercept and slope bias.
- MGCFCA results (Excel 1 sheet): containing all multigroup CFA results.
- Results calculations (Excel 2 sheet): containing all calculations for tables and figures.

Index of Supplementary Materials

Austin, B. W., & French, B. F. (2020). *Supplementary materials to: Adjusting group intercept and slope bias in predictive equations*. PsychOpen. <https://doi.org/10.23668/psycharchives.3465>

References

- Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revival of test bias research in preemployment testing. *The Journal of Applied Psychology*, 95(4), 648-680. <https://doi.org/10.1037/a0018714>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC, USA: American Educational Research Association.
- Bauer, D. J., Belzak, W. C., & Cole, V. T. (2020). Simplifying the assessment of measurement invariance over multiple background variables: Using regularized moderated nonlinear factor analysis to detect differential item functioning. *Structural Equation Modeling*, 27(1), 43-55. <https://doi.org/10.1080/10705511.2019.1642754>
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY, USA: Wiley.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5
- Church, A. T., Alvarez, J. M., Mai, N. T. Q., French, B. F., Katigbak, M. S., & Ortiz, F. A. (2011). Are cross-cultural comparisons of personality profiles meaningful? Differential item and facet

- functioning in the revised NEO personality inventory. *Journal of Personality and Social Psychology*, 101(5), 1068-1089. <https://doi.org/10.1037/a0025290>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY, USA: Holt, Rinehart and Winston.
- Ferne, T., & Rupp, A. (2007). A synthesis of 15 years of research on DIF in language testing: Methodological advances, challenges, and recommendations. *Language Assessment Quarterly*, 4(2), 113-148. <https://doi.org/10.1080/15434300701375923>
- French, B. F., & Finch, W. H. (2008). Multigroup confirmatory factor analysis: Locating the invariant referent sets. *Structural Equation Modeling*, 15(1), 96-113. <https://doi.org/10.1080/10705510701758349>
- Hancock, G. R. (1997). Structural equation modeling methods of hypothesis testing of latent variable means. *Measurement & Evaluation in Counseling & Development*, 30(2), 91-105. <https://doi.org/10.1080/07481756.1997.12068926>
- Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling*, 24(4), 524-544. <https://doi.org/10.1080/10705511.2017.1304822>
- Kuha, J., & Moustaki, I. (2015). Non-equivalence of measurement in latent variable modeling of multigroup data: A sensitivity analysis. *Psychological Methods*, 20(4), 523-536. <https://doi.org/10.1037/met0000031>
- Lomazzi, V. (2018). Using alignment optimization to test the measurement invariance of gender role attitudes in 59 countries. *Methods, data, analyses*, 12(1), 77-103. <https://doi.org/10.12758/mda.2017.09>
- McCrae, R. R., Costa, P. T., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO Personality Inventory. *Journal of Personality Assessment*, 84(3), 261-270. https://doi.org/10.1207/s15327752jpa8403_05
- Millsap, R. E. (1997). Invariance in measurement and prediction: Their relationship in the single-factor case. *Psychological Methods*, 2(3), 248-260. <https://doi.org/10.1037/1082-989X.2.3.248>
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY, USA: Routledge.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's guide* (7th Ed.). Los Angeles, CA, USA: Muthén & Muthén.
- Organisation for Economic Co-operation and Development. (2014). *PISA 2012 technical report*. Paris, France: OECD publishing. Retrieved from <http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>
- Pokropek, A., Davidov, E., & Schmidt, P. (2019). A Monte Carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance. *Structural Equation Modeling*, 26(5), 724-744. <https://doi.org/10.1080/10705511.2018.1561293>
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31-57. <https://doi.org/10.1177/0013164413498257>

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *The Journal of Consumer Research*, 25(1), 78-90.
<https://doi.org/10.1086/209528>



Methodology is the official journal of the European Association of Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology Information (ZPID), Germany.