Supplemental Table 6a: Measurement Invariance Analyses Narrative

Our measurement invariance analyses followed the strategies of Widman & Reise (1997) and Widaman and Olivera-Aguilar (in press). These methods were applied to each of the four major instruments, to investigate measurement invariance for the preferred models as indicated in the manuscript. The following narrative explicates the processing steps used and the rationale for decisions made in the modeling. Further details of the models and fit statistics are provided in Supplemental Table 6b: Factorial Invariance Models and Fit Statistics.

WAIS-IV

Factor invariance analyses. The optimal model for the WAIS-IV had a similar form in both the NNN and Pearson samples, as shown in Figure 1. Thus, we started with this model when performing factor invariance analyses to examine the degree of invariance of factor model parameter estimates across the NNN and Pearson samples. To implement these analyses, we followed standard steps to evaluate models with increasing cross-group constraints, as discussed by Widaman and Reise (1997) and Widaman and Olivera-Aguilar (in press). The initial model, termed the configural invariance model, had identical patterns of variables loading on factors, with the Verbal Comprehension and Perceptual Reasoning factors each having three indicators, the Working Memory and Processing Speed factors each having two indicators, and the second-order General factor having loadings from the four first-order factors (cf. Figure 1). To identify this model, we fixed all factor means to zero and factor variances to unity in the Pearson sample, which served as the reference group. Then, the first factor loading and the first intercept on each factor were constrained to invariance across groups to identify the factors in the NNN sample model (see Widaman & Olivera-Aguilar, in press, for additional detail and justification).

The configural invariance model had rejectable statistical fit to the data, $\chi^2(60) = 372.72$, $p < .0001$, as expected, given the large sample size. But the TLI of .972 and RMSEA of .050 both indicated close model fit to the data, and the BIC of 152531.38 served as a reference value for this series of models. The second model, for weak factorial invariance, constrained the remaining factor loadings to invariance across groups. Although this model was associated with a significant increase (i.e., worsening) in model fit, with $\Delta\chi^2(10) = 56.25$, $p < .0001$, the TLI and RMSEA were unchanged, and the lowered BIC value of 152504.41 indicated improved fit of the model to data.

The third model was the strong factorial invariance model, which added cross-group constraints on intercepts to the weak factorial invariance model. The strong invariance model led to a rather large increase in the statistical index of model misfit, $\Delta\chi^2(6) = 229.66$, $p < .0001$, the TLI (.958) and RMSEA (.061) both worsened by more than .01, and the increased BIC value of 152684.14 all indicated much poorer fit of the model to data. Modification indices indicated that two intercepts were rather different across samples, the intercepts for the Information and Block Design indicators. When invariance constraints were freed on these two intercepts, the resulting partial strong factorial invariance model displayed much improved fit to the data, $\chi^2(74) = 486.87$, $p < .0001$, and the TLI (.970) and RMSEA (.052) were also considerably improved.

The fourth model added invariance constraints across groups on unique factor variances to the strong invariance model and is termed the strict factorial invariance model. We continued to allow the intercepts for the Information and Block Design indicators to vary across groups, but imposed cross-group constraints on all 10 unique variances. Given the non-invariance of two intercepts, this model should be identified as a partial strict factorial invariance model. This model displayed non-significant loss in fit relative to the partial strong invariance model, $\Delta\chi^2(10) = 13.43$, $p = .20$, and the TLI of .973, RMSEA of .049, and BIC of 152459.23 were all the best of any models considered thus far.

Given the close fit of the partial strict factorial invariance model, we fit two additional models. The first of these models added cross-group constraints on the factor variances. Although the index of statistical fit worsened significantly, $\Delta\chi^2(5) = 18.36$, $p = .003$, the TLI of .974, RMSEA of .048, and BIC of 152.435.99 all exhibited improved fit of the model to data. The final model additionally imposed cross-group invariance constraints on factor means. This model showed very large worsening of statistical fit, $\Delta\chi^2(4) = 497.23$, $p < .0001$, and the TLI of .946, RMSEA of .069, and BIC of 152899.93 were all the worst of any of the models.

Group differences in factor mean levels are of notable importance in these WAIS-IV analyses, as these were the only model parameter estimates that differed across groups. Because of the way in which models were identified – with factor means of zero and variances of 1.0 in the Pearson sample, mean differences for the NNN sample were in a Cohen's *d* metric. Relative to the Pearson sample, the NNN sample had a mean on the General factor that was about one-third of a *SD* lower, $M = -0.326$ ($SE = 0.036$). Interestingly, the NNN sample exhibited little mean difference from the Pearson sample on the Verbal Comprehension, $M = 0.027$ ($SE = .041$), and Perceptual Reasoning factors, $M = -0.118$ ($SE = 0.044$). However, relative to the Pearson sample, the NNN sample had mean levels on the Working Memory and Processing Speed factors that were substantially lower, more than a half-*SD* in magnitude lower, with $M = -0.598$ ($SE = 0.043$) and $M = -0.615$ ($SE = 0.046$), respectively.

WMS-IV

Factor invariance analyses. The best fit models for the WMS-IV also had a similar form in both the NNN and Pearson samples, as shown in Figure 2. Factor invariance analyses for the WMS-IV followed the same steps as outline above in analyses of the WAIS-IV. Thus, we started with a configural invariance model that had minimal constraints, specifically only the constraints required to identify the model statistically in both samples. As shown in Figure 2, each of the latent variables in these analyses had three indicators. Because the distributions of the recognition manifest variables were strongly negatively skewed and had a ceiling effect, these three variables were identified as censored above, and the recommended WLSMV method of estimation was used. Unfortunately, the BIC is not available under WLSMV estimation, wo we sued the SRMR as an additional index of fit, with values < .08 indicating close model fit. In addition, given the constraints to identify the model, differences in the χ^2 index of fit could not be estimated, so only overall model fit values are reported.

The configural invariance model had rejectable statistical fit to the data, $\chi^2(36) = 76.41$, $p < .0001$, as expected, given the large sample size. But the TLI of .981 and RMSEA of .030 both indicated close model fit to the data, and the SRMR was very small, at .021. The second model, for weak factorial invariance, constrained the remaining factor loadings to invariance across groups. This model had a modest increase (i.e., worsening) in model fit, with model $\chi^2(46) = 100.65$, $p < .0001$, but the TLI of .979 and RMSEA of .031 were essentially unchanged, and the SRMR increased only slightly, to .034.

The third model was the strong factorial invariance model, which added cross-group constraints on intercepts to the weak factorial invariance model. The strong invariance model had fairly good overall model fit, $\chi^2(51) = 110.81$, $p < .0001$, and all three measures of practical fit – the TLI (.980), RMSEA (.030), and SRMR (.035) – were essentially unchanged, so model fit was deemed close

The fourth model added invariance constraints across groups on unique factor variances to the strong invariance model and is termed the strict factorial invariance model. This model displayed only slightly worse overall fit relative to the strong invariance model, $\chi^2(62) = 134.39$, $p < .0001$, and the TLI of .980 and RMSEA of .030 were unchanged, and the SRMR increased only slightly to .053, but all indices indicated close model fit to the data..

Given the close fit of the strict factorial invariance model, we fit two additional models. The first of these models added cross-group constraints on the factor variances. The index of statistical fit worsened considerably, $\chi^2(68) = 530.81$, $p < .0001$. Moreover, the TLI of .882, RMSEA of .073, and SRMR of .249 all indicated very poor fit of the model to data. The final model additionally imposed cross-group invariance constraints on factor means. This model also showed much worse statistical fit, $\Delta\chi^2(72) = 1333.47$, $p < .0001$, and the TLI of .697, RMSEA of .118, and SRMR of .386 were all the worst of any of the models.

Group differences in factor means and factor variances across groups were present in these WMS-IV analyses. Given limitations of space, we present here only the group differences in mean levels; differences in factor variance are shown in supplemental material. Because of the way in which models were identified – with factor means of zero and variances of 1.0 in the Pearson sample, mean differences for the NNN sample were in a Cohen's *d* metric. Relative to the Pearson sample, the NNN sample had a large mean difference on the General factor that was about one full *SD* in magnitude, $M = -0.986$ ($SE = 0.075$). The NNN sample exhibited similar large differences on the first-order Visual Reproduction, $M = -1.077$ ($SE = .096$), Logical Memory, $M = -0.897$ ($SE = .085$), and Verbal Paired Associate factors, $M = -0.983$ ($SE = .126$). For the Recognition-Familiarity factor, given unequal variances, we set the mean of the NNN sample to zero and variance to 1.0, and relative to this the Pearson sample mean was M = 1.026, SE = .061)

CVLT3

Factor invariance analyses. As found for the preceding instruments, the optimal model for the CVLT3 had a similar form in both the NNN and Pearson samples, as shown in Figure 3. Once again, we followed the same modeling steps in our invariance analyses as we had used for the WAIS-IV and WMS-IV. Thus, the configural invariance model for the CVLT3 had the basic form as that shown in Figure 3 for both the NNN and Pearson samples, a model with a single second-order factor and four first-order factors.

The configural invariance model had rejectable statistical fit to the data, $\chi^2(116) = 547.30$, $p < .0001$, as expected, given the large sample size. But the TLI of .954 and RMSEA of .074 were in an acceptable range, and the BIC of 77584.65 served as a reference value for these CVLT3 models. The second model, weak factorial invariance, constrained the remaining factor loadings to invariance across groups. This model led to a non-significant increase (i.e., worsening) in model fit, with $\Delta\chi^2(12) = 13.28$, $p = .35$, and the TLI of .959, RMSEA of .071, and BIC value of 77511.39 all indicated improved fit of the model to data.

The third model was the strong factorial invariance model, which added cross-group constraints on intercepts to the weak factorial invariance model. The strong invariance model led to a moderate increase in the statistical index of model misfit, $\Delta\chi^2(9) = 81.50$, $p < .0001$, but the TLI (.955) and RMSEA (.074) worsened only slightly, so no model modifications were pursued.

The fourth model added invariance constraints across groups on unique factor variances to the strong invariance model and is termed the strict factorial invariance model. This model displayed a significant loss in statistical fit relative to the strong invariance model, $\Delta\chi^2(16) = 39.77$, $p < .0001$, but the TLI of .958, RMSEA of .071, and BIC of 77452.37 all displayed improved levels of fit.

Given the close fit of the strict factorial invariance model, we fit two additional models. The first of these models added cross-group constraints on the factor variances. Although the index of statistical fit worsened significantly, $\Delta\chi^2(5) = 18.99$, $p = .002$, the TLI of .958 and RMSEA of .071 were unchanged, and BIC of 77435.30 was the best of any of the models. The final model additionally imposed cross-group invariance constraints on factor means. This model showed relatively large worsening of statistical fit, $\Delta\chi^2(4) = 72.47$, $p < .0001$, and the TLI of .954, RMSEA of .075, and BIC of 77478.92 all had somewhat poor fit.

Group differences in factor mean levels are of notable importance in these CVLT3 analyses, as these were the only model parameter estimates that differed substantially across groups. Because of the way in which models were identified – with factor means of zero and variances of 1.0 in the Pearson sample, mean differences for the NNN sample were in a Cohen's *d* metric. Relative to the Pearson sample, the NNN sample had a mean on the General factor that was about one-third of a *SD* lower, $M = -0.337$ ($SE = 0.059$). The NNN sample had a moderatly large mean difference on the Attention Span factor, $M = -0.583$ ($SE = 0.078$). The NNN sample exhibited relatively small mean difference from the Pearson sample on the Learning Efficiency, $M = -0.181$ ($SE = .061$), and Delayed Memory factors, $M = -0.200$ ($SE = 0.057$). Relative to the Pearson sample, the NNN sample had a mean level on the Inaccurate Memory that was rather lower, over than one-third-*SD* in magnitude, with $M = -0.384$ ($SE = 0.075$).

D-KEFS

Factor invariance analyses. As with preceding instruments, the optimal model for the D-KEFS had a similar form in both the NNN and Pearson samples, as shown in Figure 4. The model has a second-order General factor, three first-order factors that are test-based (for Color Word Interference, Trail Making, and Fluency), and a fourth first-order factor for Inhibition/Switching that has at least one indicator from each test.

The configural invariance model had rejectable statistical fit to the data, $\chi^2(90) = 179.47$, $p < .0001$, as expected, given the large sample size. But, the TLI of .974 and RMSEA of .037 both indicated close model fit to the data, and the BIC of 66231.63 served as a reference value for this series of models. The second model, for weak factorial invariance, constrained the remaining factor loadings to invariance across groups. Although this model was associated with a significant increase (i.e., worsening) in model fit, with $\Delta\chi^2(15) = 32.33$, $p = .006$, the TLI (.974) and RMSEA (.038) were essentially unchanged, and the lowered BIC value of 66155.04 indicated improved fit of the model to data.

The third model was the strong factorial invariance model, which added cross-group constraints on intercepts to the weak factorial invariance model. The strong invariance model led to a rather large increase in the statistical index of model misfit, $\Delta\chi^2(8) = 66.26$, $p < .0001$, the TLI (.962) and RMSEA (.045) both worsened noticeably, and the increased BIC value of 66163.20 all indicated poorer fit of the model to data. Modification indices indicated that the intercept for the Motor Speed indicator was responsible for the worsened model fit. When the invariance constraints were freed on this intercept, the resulting partial strong factorial invariance model displayed much improved fit to the data, $\chi^2(112) = 242.94$, $p < .0001$, and the TLI (.970), RMSEA (.041), and BIC (66135.34) all improved.

The fourth model added invariance constraints across groups on unique factor variances to the strong invariance model, leading to the strict factorial invariance model. We continued to allow the intercept for Motor Speed to vary across groups but imposed cross-group constraints on all 12 unique variances. This initial partial strict factorial invariance model had a rather large increase in statistical fit relative to the partial strong invariance model, $\Delta\chi^2(12) = 39.75$, $p < .0001$. In this model, it appeared that the unique variance for the Motor Speed indicator was responsible for most of the worsened fit. So, relaxing the invariance constraint on the Motor Speed unique variance led to a model with improved overall fit, $\chi^2(123) = 258.01$, $p < .0001$, and the TLI (.972), RMSEA (.039), and BIC (66070.53) all improved.

Given the close fit of the partial strict factorial invariance model, we fit two additional models. The first of these models added cross-group constraints on the factor variances. The index of statistical fit worsened significantly, $\Delta\chi^2(5) = 99.51$, $p < .0001$, and the TLI of .955, RMSEA of .050, and BIC of 66133.74 all exhibited clearly worsened fit of the model to data. The final model additionally imposed cross-group invariance constraints on factor means. This model also showed a relatively large worsening of statistical fit, $\Delta\chi^2(4) = 94.53$, $p < .0001$, and the TLI of .937, RMSEA of .058, and BIC of 66199.22 were all in the unacceptable range.

Group differences in factor means and factor variances were in evidence in these D-KEFS models. Given constraints of space, we discuss here only the mean level differences. Relative to the Pearson sample, the NNN sample had a mean on the General factor that was over one-half of a *SD* lower, $M = -0.569$ ($SE = 0.098$). The NNN had its lowest mean level on a first-order factor on the Color-Word Interference factor, with $M = -0.906$ ($SE = 0.108$). The NNN sample exhibited smaller, but still fairly large mean difference from the Pearson sample on the Trail Making, $M = -0.594$ ($SE = .147$), Fluency, $M = -0.316$ ($SE = .109$), and Perceptual Reasoning factors, $M = -0.459$ ($SE = 0.135$).