**Testing Measurement Invariance**

The same steps for testing longitudinal invariance were used for our three study variables. First, the latent variable indicators (i.e., the individual items) at baseline were examined for standardized loadings of less than .35. These indicators were removed from the latent variable given their low factor loadings. For study variables with ten or more indicators, parcels of indicators (see Little et al., 2022) were created by randomly assigning (via random number generator) indicators to parcels, which were used as indicators of the latent factor of interest. For models that do not meet the fit criteria (see Whittaker, 2011), modification indices were examined for patterns of suggested fit increase. The revised factor model was examined for goodness of fit and conceptual clarity (i.e., indicators loading positively on the latent factor).

This initial model (created after following the above steps) was used to test longitudinal invariance for each study variable. Testing longitudinal invariance consisted of creating four models, each increasingly constrained in the number of parameters set to be equivalent across time. In the *configural* model, the number of factors and patterns of zero and non-zero factor loadings are constrained to be identical over time and all other parameters in the model are allowed to vary freely. If configural invariance is met, the factors cannot be assumed to be the same construct over time and may not lie on the same scale. As such, the configural model is not sufficient to assume longitudinal invariance, but instead is used as the baseline comparison model for the subsequent models. Next fit to the data is the *metric* model, also known as the *weak* invariance model. The metric model further constrains the configural model by constraining items to load equally on the latent factors across time. In other words, not only do all factors have the same number of items, but the factor loadings on item 1, for example, is held constant over time, similarly for item 2, item 3, etc. The remaining parameters, including the observed variable means, are allowed to vary freely across timepoints and thus we cannot make longitudinal assumptions of change from differing latent variable means.

The next level of invariance is tested by the *scalar* (or strong) invariance model. This model further constrains the metric model by constraining the observed variable means to be equal over time (not just the loadings of that item on the factor). In the scalar invariance model, changes in the observed variables are due to changes in the underlying factors and that the factors lie on the same scale. Finally, the *strict* invariance model adds to the scalar model by further constraining the unique variances across time. While strict invariance is the most stringent form of invariance, it is generally considered acceptable to assess change over time if the conditions of scalar longitudinal invariance are met (Grimm et al., 2017).

These four models were built for each of the study variables, and the model fit of each were compared using the root mean squared error of approximation (RMSEA; Stiger, 1990), the Comparative Fit Index (CFI; Bentler, 1990), the Tucker-Lewis Index (TFI; Tucker & Lewis, 1973), and the standardized root square mean residual (SRMR; Maydeu-Olivares, 2017). Care will be taken to examine the relative fit of each model compared to the more-constrained version (i.e., comparing the configural model to the metric model) and to retain the model that is the most parsimonious while also maintaining appropriate fit to the data.

***Family Conflict***

The initial model of the family conflict latent variable is shown in Supplemental Table 1. This model, including all nine indicators, fit marginally adequately to the data (RMSEA = .052; CFI = .899, TLI = .866, SRMR = .035). The modification indices suggested major improvements in model fit to the data when items describing physical violence (i.e., *We get so angry we throw things; Sometimes we hit each other*) were allowed to freely covary. Additionally, major improvements to model fit were suggested by allowing the reversed-phraseditems (e.g*., We hardly ever lose our temper*) to covary as well. Each modification was added one at a time and updated modification indices were requested in order to make only the changes necessary to improve model fit. In this model (shown in Supplemental Table 2; RMSEA = .028, CFI = .978, TLI = .960, SRMR = .016), two items had factor loadings less than .35. However, removal of these items reduced model fit below standard criterion (see Supplemental Table 3) and thus they were retained.

Supplemental Table 4 shows the longitudinal invariance testing process over four models, beginning with the modified model shown in Supplemental Table 2. Model fit degraded somewhat but the metric and scalar models continued to show strong fit even with increasingly constrained parameters. When testing strict invariance, the CFI and TLI degraded below acceptable levels and thus strict invariance did not hold. Scalar invariance, however, is sufficient to proceed with longitudinal modeling (Grimm et al., 2017) and this model was retained for further analyses.

***Externalizing Behavior***

The initial model for youth externalizing behavior is shown in Supplemental Table 5. This model included all 33 items on the CBCL that are normally entered into the externalizing factor. Initial model fit was not acceptable (*c*2 (495)= 8789.374, RMSEA = .038, CFI = .776, TLI = .761, SRMR = .050). Five items had loadings under .35 and were rarely occurring in the sample, and thus were removed before the creation of item parcels. The remaining 28 items were randomly assigned to one of five parcels (i.e., three parcels of six items and two parcels of five items) and averaged to create parceled indicators. These parcels were used as indicators in a confirmatory factor analysis (CFA) for externalizing behavior at baseline and model fit was excellent (*c*2 (5) = 59.105, RMSEA = .030, CFI = .998, TLI = .996, SRMR = .007; Supplemental Table 6). Supplemental Table 7 shows the LMI process over four models, beginning with the modified, parceled model shown in Supplemental Table 6. Model fit degraded with increasingly strict invariance constraints, but the strict invariance model maintained good overall fit (*c*2 (173) = 767.761, RMSEA = .017, CFI = .994, TLI = .993, SRMR = .019).

***Internalizing Behavior***

The initial model for youth internalizing behavior is shown in Supplemental Table 8. This model included all 31 items on the CBCL that are normally entered into the internalizing factor. Initial model fit was unacceptable (*c*2 (434) = 11182.380, RMSEA = .046, CFI = .727, TLI = .707, SRMR = .054) and indicators were examined for low factor loadings (i.e., < .35). Six items had loadings under .35 and were removed before the creation of item parcels. The remaining 25 items were randomly assigned to one of five parcels (of five items each) and averaged to create parceled indicators. The parcels were used as indicators in a CFA for internalizing behavior at baseline and model fit was excellent (*c*2 (5) = 41.353, RMSEA = .025, CFI = .998, TLI = .996, SRMR = .007; Supplemental Table 9). Supplemental Table 10 shows the LMI process over four models, beginning with the modified, parceled model shown in Supplemental Table 9. Model fit degraded with increasingly strict invariance constraints, but the strict invariance model maintained good overall fit (*c*2 (57) = 767.761, RMSEA = .017, CFI = .994, TLI = .993, SRMR = .016).

Table 1

*Initial Confirmatory Factor Analysis of Family Conflict at Baseline*

Indicator β

1. We fight a lot (not physically). **.616**
2. We rarely become openly angry. **.405**
3. We get so angry that we throw things. **.448**
4. We hardly ever lose our temper. **.491**
5. We often criticize each other. **.467**
6. Sometimes we hit each other. **.547**
7. When we disagree, we try to smooth things over. .299
8. We often try to one-up or out-do one each other. **.402**
9. Raising our voice will not get us anywhere. .243

*Note. c*2 (27) = 888.738, RMSEA = .052, CFI = .899, TLI = .866, SRMR = .035. Bolded coefficients are statistically significant and greater than or larger than .35.

Table 2

*Modified CFA of Family Conflict at Baseline: Covariances added*

Indicator β

1. We fight a lot (not physically). **.643**
2. We rarely become openly angry. **.355**
3. We get so angry that we throw things. **.409**
4. We hardly ever lose our temper. **.442**
5. We often criticize each other. **.485**
6. Sometimes we hit each other. **.529**
7. When we disagree, we try to smooth things over. .277
8. We often try to one-up or out-do one each other. **.419**
9. Raising our voice will not get us anywhere. .196

*Note. c*2 (20) = 210.311, RMSEA = .028, CFI = .978, TLI = .960, SRMR = .016. Bolded coefficients are statistically significant and greater than or larger than .35.

Table 3

*Modified CFA of Family Conflict: Low loadings removed*

Indicator β

1. We fight a lot (not physically). **.644**
2. We rarely become openly angry. **.355**
3. We get so angry that we throw things. **.410**
4. We hardly ever lose our temper. **.442**
5. We often criticize each other. **.482**
6. Sometimes we hit each other. **.529**
7. When we disagree, we try to smooth things over. ---
8. We often try to one-up or out-do one each other. **.421**
9. Raising our voice will not get us anywhere. ---

*Note. c*2 (27) = 1345.168, RMSEA = .064, CFI = .846, TLI = .795, SRMR = .072. Bolded coefficients are statistically significant and greater than or larger than .35.

Table 4

*Tests of Longitudinal Measurement Invariance of Family Conflict*

Configural Metric Scalar Strict

P 196 172 148 120

*c*2 (*df*) 2555.495(506) 2923.940(530) 4440.674(554) 6815.194(582)

AIC 320745.892 321152.965 322901.532 325752.537

BIC 322192.587 322422.513 323993.934 326638.268

RMSEA .018 .020 .024 .030

CFI .959 .953 .923 .877

TLI .950 .944 .913 .867

SRMR .024 .027 .030 .041

*Note.* All listed chi-square values are statistically significant at *p* < .05. P = parameters. df = degrees of freedom. AIC = Akaike Information Criteria. BIC = Bayesian Information Criteria. RMSEA = root mean square error of approximation. CFI = confirmatory fit index. TLI = Tucker-Lewis index. SRMR = standardized root mean residual.

Table 5

*Initial Confirmatory Factor Analysis of Externalizing at Baseline*

Indicator β

1. Argues **.650**
2. Brags **.441**
3. Mean **.548**
4. Demands attention **.585**
5. Destroys own belongings **.569**
6. Destroys others’ belongings **.626**
7. Disobeys at home **.658**
8. Disobeys at school **.558**
9. No guilt **.559**
10. Easily jealous **.537**
11. Fights **.486**
12. Bad companion **.386**
13. Lies or cheats **.580**
14. Attacks **.519**
15. Prefers older peers **.344**
16. Runs away .324
17. Screams **.590**
18. Sets fires .170
19. Shows off **.491**
20. Steals from home **.448**
21. Steals from others?? **.369**
22. Stubborn **.610**
23. Mood changes **.587**
24. Swears **.461**
25. Talks too much **.410**
26. Teases **.524**
27. Temper **.652**
28. Thinks about sex .221
29. Threatens **.528**
30. Truant .147
31. Loud **.507**
32. Alcohol or drugs .014
33. Vandalism **.345**

*Note. c*2 (495) = 8789.375, RMSEA = .038, CFI = .776, TLI = .761, SRMR = .050. Bolded coefficients are statistically significant and greater than or larger than .35.

Table 6

*Modified, Parceled CFA of Externalizing Behavior at Baseline*

 β

Parcel 1 .**831**

*Temper*

*Shows off*

*Fights*

*Attacks*

*Bad companion*

*Steals from home*

Parcel 2 .**819**

*Demands attention*

*Loud*

*Brags*

*Threatens*

*Vandalizing*

*No guilt*

Parcel 3 .**848**

*Swears*

*Disobeys at home*

*Screams*

*Argues*

*Destroys own belongings*

*Steals from others*

Parcel 4 .**825**

*Teases*

*Talks too much*

*Mood changes*

*Lies or cheats*

*Disobeys at school*

Parcel 5 .**820**

*Destroys others’ things*

*Mean*

*Jealous*

*Prefers older peers*

*Stubborn*

*Note. c*2 (5) = 59.105, RMSEA = .030, CFI = .998, TLI = .996, SRMR = .007. Bolded coefficients are statistically significant and greater than or larger than .35.

Table 7

*Tests of Longitudinal Measurement Invariance of Child Externalizing Behavior*

Configural Metric Scalar Strict

P 96 84 72 57

*c*2 (*df*) 388.755(134) 466.887(146) 680.051(158) 995.355(173)

AIC -191864.916 -191731.556 -191415.217 -190867.332

BIC -191156.298 -191111.516 -190883.664 -190446.591

RMSEA .013 .014 .017 .020

CFI .998 .997 .995 .993

TLI .997 .996 .994 .992

SRMR .012 .014 .015 .019

*Note.* All listed chi-square values are statistically significant at *p* < .05. P = parameters. df = degrees of freedom. AIC = Akaike Information Criteria. BIC = Bayesian Information Criteria. RMSEA = root mean square error of approximation. CFI = confirmatory fit index. TLI = Tucker-Lewis index. SRMR = standardized root mean residual.

Table 8

*Initial CFA of Child Internalizing Behavior*

Indicator β

1. Lonely **.508**
2. Cries **.432**
3. Fears do bad **.435**
4. Perfect **.437**
5. Unloved **.540**
6. Out to get **.443**
7. Worthless **.584**
8. Rather be alone **.411**
9. Nervous **.570**
10. Fearful **.589**
11. Dizzy .324
12. Guilty **.472**
13. Tired **.383**
14. Aches .303
15. Headaches .322
16. Nausea **.382**
17. Eye problems .185
18. Skin problems .215
19. Stomach problems **.375**
20. Vomit .183
21. Won’t talk **.407**
22. Secretive **.441**
23. Shy **.570**
24. Self conscious **.417**
25. Stares **.386**
26. Sulks **.551**
27. Suspicious **.389**
28. Under-active **.405**
29. Sad **.601**
30. Withdrawn **.463**
31. Worries **.606**

*Note. c*2 (434) = 11182.380, RMSEA = .046, CFI = .727, TLI = .707, SRMR = .054. Bolded coefficients are statistically significant and greater than or larger than .35.

Table 9

*Modified, Parceled CFA of Internalizing Behavior at Baseline*

 β

Parcel 1 **.835**

*Secretive*

*Out to get*

*Worries*

*Under-active*

*Worthless*

Parcel 2 .**795**

*Sad*

*Nauseous*

*Self conscious*

*Unloved*

*Shy*

Parcel 3 **.689**

*Rather be alone*

*Stomach aches*

*Cries*

*Suspicious*

*Stares*

Parcel 4 **.753**

*Lonely*

*Won’t talk*

*Perfect*

*Withdrawn*

*Fear do bad*

Parcel 5 **.780**

*Guilty*

*Sulks*

*Tired*

*Fearful*

*Nervous*

*Note. c*2 (5) = 41.353, RMSEA = .025, CFI = .998, TLI = .996, SRMR = .007. Bolded coefficients are statistically significant and greater than or larger than .35.

Table 10

*Tests of Longitudinal Measurement Invariance of Child Internalizing Behavior*

Configural Metric Scalar Strict

P 96 84 72 57

*c*2 (*df*) 317.245(134) 438.462(146) 679.159(158) 767.761(173)

AIC -155665.535 -155470.952 -155124.972 -154982.170

BIC -154956.917 -154850.912 -154593.508 -154561.428

RMSEA .011 .013 .017 .017

CFI .998 .997 .994 .994

TLI .997 .996 .993 .993

SRMR .010 .014 .014 .016

*Note.* All listed chi-square values are statistically significant at *p* < .05. P = parameters. df = degrees of freedom. AIC = Akaike Information Criteria. BIC = Bayesian Information Criteria. RMSEA = root mean square error of approximation. CFI = confirmatory fit index. TLI = Tucker-Lewis index. SRMR = standardized root mean residual.