**Areas of local ill fit for the Pitt-2 model.** In examining the standardized residuals (SRs), we considered residuals greater than the absolute value of 2.58, which corresponds to a *p* value of .01, to be significant because of the large sample size (Brown, 2006). Across the nine Pitt-2 models, there were areas of local ill fit in each one. With 17 standardized residuals in each model, there were 3 significant residuals for age 4 males, with the model underestimating the “get even/defies” relationship (SR = 2.93) and overestimating the “defies/temper” relationship (-2.98) and the” angry/argues relationship” (-3.15). For 5-year-old boys, there were also 3 significant SRs, with the model underestimating the get even/defies (SR = 3.60) relationship and overestimating the defies/temper relationship (-4.23) and the argues/touchy relationship (-2.96). For 6-year-old boys, there were 2 significant SRs, with the model underestimating the get even/defies (SR = 3.79) relationship and the get even/temper relationship (2.89).

For 4-year-old girls, there was 1 significant SR, with the model underestimating the get even/defies (SR = 3.44) relationship. For 5-year-old girls, there were 3 significant SRs, with the model underestimating the get even/defies (SR = 5.05) relationship and overestimating the argues/touchy (-2.96) relationship and the angry/argues relationship (-3.54). For 6-year-old girls, there were also 3 significant SRs, with the model underestimating the angry/touchy (SR = 3.52) relationship and overestimating the defies/temper (-3.31) relationship and the angry/argues (-3.52) relationship.

For the 4-year-old combined sex groups, there were 4 significant SRs, with the model underestimating the get even/defies (SR = 4.48) relationship and the get even/temper (2.74) relationship, and overestimating the defies/temper (-3.32) relationship and the angry/argues (-3.45) relationship. For the 5-year-old combined sex groups, there were also 4 significant SRs, with the model underestimating the get even/defies (SR = 5.89) relationship and the argues/defies (11.54) relationship, and overestimating the argues/touchy (-4.06) relationship and the angry/argues (-3.41) relationship. For the 6-year-old combined sex group, there were 4 significant SRs, with the model underestimating the get even/defies (4.06) relationship and the get even/temper (SR = 3.12) relationship and overestimating the defies/temper (-3.57) relationship and the angry/argues (-4.08) relationship. While no combination of factors showed significant residuals in all 9 groups, the most common problems involved get even with defies (8 of 9 models), angry/argues (6 of 9 models), and defies/temper (5 of 9 models). These areas of problematic local fit indicate that, while the Pitt-2 model showed the better overall fit compared to the alternative models, there is still room for improvement for that model.

Examining modification indices may provide clues about ways in which the models could be improved. Modification indices (MIs) can provide suggestions about specific paths that might be added to a model to improve model fit. MIs *>* 3.84 could possibly improve a model at a statistically significant level (p < .05). Modification indices, however, are sensitive to sample size—it is possible that estimating the parameter associated with a significant MI could result in a factor loading that is very small and of little value. For these reasons, examining MIs to gain insight into areas of poor model fit should also include examination of completely standardized expected parameter change (EPC) scores.

For the 3 models for boys, there were 6 significant MIs. Since desirable factors loadings have values of .70 or greater, the completely standardized EPCs for these factor loadings for 4 were relatively small and in some instances were negative (age 5: ODDNA🡺temper, EPC = .20; ODDNA🡺 argues, EPC = -.19; ODDB 🡺 get even, EPC = .18; age 6 ODDNA 🡺 angry, -.58) or moderate in magnitude (ODDB 🡺 get even, EPC = .40). One was large, age 4 ODDNA🡺 temper, EPC =1.14.

For the girl models, there were 11 significant MIs. There were xxx with small or negative EPCs (age 4 ODDNA 🡺 argues, EPC = .13; age 4 ODDB = angry, EPC = -.89; age 5 ODDNA 🡺 argues, EPC = -.24, age 4 ODDNA 🡺 defies, EPC = .28; ODDB 🡺 defies, EPC = .28; age 5 ODDB 🡺 touchy, EPC = 0.33; age 6 ODDNA 🡺 temper, EPC = .13; ODDNA 🡺 argues, EPC = -.14; age 6 ODDB 🡺 angry, EPC = .24; age 6 ODDB 🡺 get even, EPC = .29) and 2 were moderate (age 4 ODDB 🡺 gets even, EPC = .50; age 5 ODDB 🡺 get even, EPC = .42).

For the combined sex models, there were 10 significant MIs, with most exhibiting small or negative EPCs (age 4 ODDNA 🡺 temper, EPC = .15; ODDNA 🡺 argues, EPC = -.19; ODDB 🡺 angry, EPC = -.19; ODDNA 🡺 temper, EPC = .15; age 5 ODDNA 🡺 temper, EPC = .10; age 5 ODDNA 🡺 argues, EPC = -.17; ODDB = touchy, EPC = -.24; ODDB 🡺 get even, .25; age 6 ODDNA 🡺 temper, EPC = .11; ODDNA 🡺 argues, ECP = -.17; ODDB 🡺 angry, EPC = -.40) and one was moderate (age 6 ODDB 🡺 get even, EPC = .38). See table xxx in on-lie supplementary tables for MI values.

These results suggest that adding the factor loadings associated with these MIs would improve model fit overall for all or most of those models. However, the low or moderate factor loadings would be eliminated because they were far below the desired standard for factor loadings of .70. The one large factor loading (age 4 boys, ODDNA 🡺 temper) might be retained, but doing so would have the disadvantage of eliminating configural invariance for boys with the Pitt-2 model. For these reasons, adding that cross-factor would not be desirable.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Pitt two-factor model: standardized residuals** | | |  |  |
| **Group** | # nonsignificant standardized residuals | # significant standardized residuals | Model underestimates (positive residuals) | Model overestimates (negative) |
| Males |  |  |  |  |
| Age 4 | 14 | 3 | get even/defies (2.93) | defies/temper (-2.98); angry/argues (-3.15) |
| Age 5 | 14 | 3 | get even/defies (3.60) | defies/temper (-4.23); argues/touchy (-2.96) |
| Age 6 | 15 | 2 | get even/defies (3.79); get even/temper (2.89) | None |
| Females |  |  |  |  |
| Age 4 | 16 | 1 | Get even/defies (3.44) | None |
| Age 5 | 14 | 3 | get even/defies (5.05) | argues/touchy (-2.96); angry/argues (-3.54) |
| Age 6 | 14 | 3 | angry/touch (3.52) | defies/temper (-3.31); angry/argues (-3.52) |
| Both genders pooled |  |  |  |  |
| Age 4 | 13 | 4 | Get even/defies (4.48) ; get even/temper (2.74) | defies/temper (-3.22); angry/argues (-3.45) |
| Age 5 | 13 | 4 | get even/defies (5.89); argues/defies (11.56) | argues/touchy (-4.06); angry/argues (-3.41) |
| Age 6 | 13 | 4 | get even/defies (4.06); get even/temper (3.12) | defies/temper (-3.57); angry/argues (-4.08) |

|  |  |
| --- | --- |
| **Significant Modification Indices (> 3.84)** | |
| Group | MI |
| Males | Paths with significant MIs (MI value/completely standardized estimate of factor loading) |
| Age 4 | ODDNA==>temper (23.81/1.14) |
| Age 5 | ODDNA==>temper (6.65/.20); ODDNA==>argues (7.30/-.19); ODDB ==>get even (5.09/.18) |
| Age 6 | ODDB==>angry (17.63/-.58); ODDB==>get even (22.03/.40) |
| Females |  |
| Age 4 | ODDNA==>argues (8.43/.13); ODDB==>angry (15.92/-.89); ODDB==> get even (16.41/.50) |
| Age 5 | ODDNA==argues (6.47/-.24); ODDNA==>defies (8.35/.28); ODDB==>touchy (3.77/-.33); ODDB==>get even (12.11/.42) |
| Age 6 | ODDNA==>temper (3.89/.13); ODDNA==> argues (5.05/-.14); ODDB ==> angry (5.78/-.24); ODDB==>get even (6.74/.29) |
| Both genders pooled |  |
| Age 4 | ODDNA ==> temper (6.01/.15); ODDNA ==> argues (10.48/-.19); ODDB ==> angry (5.86/-.19) |
| Age 5 | ODDNA ==> temper (3.74/.10); ODDNA ==> argues (13.45/-.17); ODDB ==> touchy (5.08/-.24); ODDB ==> get even (17.64/.25) |
| Age 6 | ODDNA ==> temper (5.66/.11); ODDNA ==> argues (6.96/-.12); ODDB ==> angry (22.63/-.40; ODDB ==> get even (30.94/.38) |