**Supplemental Material**

**Explanation of changes from preregistration**

A number of changes were made in preparation of this article relative to the data analysis plan in the preregistration, and are described here. Originally, we planned to use only teacher reports (TR) of bullying and victimization, and planned to use Fall, rather than Spring (T3) data. However, initial models using TR only with no parcels (8 bullying items, 4 victimization items) provided marginal fit to data [Time 1: χ2 (42) = 178.07, *p* < .0001, RMSEA = .10, CFI = .92, SRMR = .05; Time 2: χ2 (42) = 123.84, *p* < .0001, RMSEA = .10, CFI = .94, SRMR = .03]. Parcels were then created from relational bullying items using TR only to equalize the number of indicators per factor (4 bullying parcels, 4 victimization items) but these models did not converge. Therefore, RA reports on the same measures were added to increase the variance entered into the models, and these are the results presented. RA reports were not available at the Fall timepoint, so T3 data were substituted.

We initially planned to examine hostile attribution biases for relational provocations (HAB-R) specifically. However, the HAB-R subscale was not reliable, even after dropping one item (α = .64). Therefore, we collapsed across subtypes for main analyses, as specified in our preregistered contingency plan. However, as these internal consistency values are consistent with prior published work with this measure (Godleski & Ostrov, 2020; Perhamus & Ostrov, 2021), we ran additional post-hoc models examining effects with HAB-R specifically, which are presented below. Additionally, we examined age as a covariate, but it was not significantly correlated with any outcome variables (*r*s = -.08 - .11, *p*s >.05), so it was dropped from models. Room temperature during the stressor task was correlated with SCL-R (*r* = .29, *p* = .008). Models were run with temperature included as a covariate with no meaningful changes to results. Therefore, due to concerns regarding statistical power and because the strength of this correlation was below our proposed a priori cutoff of *r* = .30 for inclusion, it was not included in models.

Consistent with our preregistered plan, we first conducted path analysis models without physical forms of bullying and victimization included as covariates. However, including these variables is currently considered a best practice in the field, to account for generally high overlap across forms of aggression and victimization and ensures effects are related specifically to the form of aggression of interest (Ostrov & Kamper, 2015). Unlike our preregistered plan, we did not enter SCL-R separately from HAB in nested path models (i.e., SCL-R at step 2, HAB at step 3, interaction at step 4), due to the low correlation between SCL-R and HAB and to reduce model complexity.

Finally, in the preregistration we planned to conduct all path analyses as hierarchical regression models in SPSS. However, when conducting the bifactor analysis, several individuals were missing data on the bullying and victimization questionnaires at both time points (see main text). Because the bifactor analyses were run at one time point, missing data could not be accommodated using full information maximum likelihood (FIML) for these models. Therefore, to accommodate missing data and maximize statistical power by including all potential participants in tests of hypotheses related to HAB and SCL-R, we conducted these tests in Mplus as presented in the main text.

**HAB-R path models**

Models testing effects of HAB-R specifically are largely consistent with models including the full HAB composite (as described in main text). However, there were a few changes to note. There was significant improvement in model fit from step 1 to step 2 in models predicting change in the victimization group factor [χ2(2) =11.88, *p* = .003]. Likewise, models predicting change in the relational bullying group factor found significant improvement in model fit from both step 1 to step 2 [χ2(2) =13.45, *p* = .001] and step 2 to step 3 [χ2(1) =10.00, *p* = .002]. However, as in the HAB composite models, no significant effects of SCL-R, HAB-R, or their interaction emerged (βs = -.14 - .21, *p*s = .06 - .65). Finally, in models predicting change in the general factor, there was significant change in log-likelihood from step 1 to step 2 [χ2(2) =12.92, *p* = .002] in addition to from step 2 to step 3 [χ2(1) = 4.18, *p* = .04]. However, the direct effect of HAB-R was non-significant (β = .08, 95% CI [-.12, .27] *p* = .45), and the interaction term was marginally significant (β = .12, 95% CI [-.006, .25] *p* = .06) unlike models using the HAB composite. This is consistent with the lower reliability of the HAB-R subscale relative to the full HAB composite.