Schoolyard Politics: Measuring and Explaining Variation in State Anti-Bullying Policy Comprehensiveness

Supplementary Information

Detailed Description of Measuring State Anti-bullying Policy Comprehensiveness

This supplement walks through the decisions made in choosing an estimation strategy for measuring the latent concept of anti-bullying policy comprehensiveness. It begins by presenting summary statistics for each item included and then evaluates four options for item response theory (IRT) modeling.

In terms of summary statistics, Table S1 provides the proportions of the total number of states (n=50) that either has (1) or has not (0) adopted each policy item. They range from a low of 32 percent adoption (i.e., mental health) to a high of 94 percent adoption (i.e., scope and district policy). Table 1 from the main text provides a description of each item.

Table S1
Proportions of Each Level of Response for Each Bullying Policy Item

Item	0	1
Purpose	0.12	0.88
Scope	0.06	0.94
Prohibited behavior	0.08	0.92
Enumerated groups	0.64	0.36
District policy	0.06	0.94
District policy review	0.54	0.46
Definitions	0.36	0.64
Reporting	0.20	0.80
Investigations	0.24	0.76
Written records	0.50	0.50
Consequences	0.08	0.92
Mental health	0.68	0.32
Communications	0.12	0.88
Training/prevention	0.22	0.78
Transparency/monitoring	0.54	0.46
Legal remedies	0.60	0.40

The main text reported a fairly large Cronbach's α for all times, meaning that intercorrelations among the items are fairly high and the measure of the latent concept of comprehensiveness has decent internal consistency. Table S2 also reports the Cronbach's α when

each item is iteratively removed. All of the resulting measures of consistency fall within a narrow range between 0.84 and 0.86. Thus, removing any given component does not improve the consistency of the latent measure.

Table S2
Chronbach's α for all items and all item when one item is excluded

Item	Chronbach's α
All items	0.86
Purpose excluded	0.85
Scope excluded	0.85
Prohibited behavior excluded	0.86
Enumerated groups excluded	0.85
District policy excluded	0.86
District policy review excluded	0.85
Definitions excluded	0.85
Reporting excluded	0.85
Investigations excluded	0.84
Written records excluded	0.85
Consequences excluded	0.85
Mental health excluded	0.86
Communications excluded	0.85
Training/prevention excluded	0.85
Transparency/monitoring excluded	0.86
Legal remedies excluded	0.86

Using the ltm package (Rizopoulos 2006) in R (R Core Team 2015), I tested four different options for estimating a latent measure of comprehensiveness using IRT modeling for dichotomous data. These options depend on the treatment of the guessing, discrimination, and difficulty parameters of the model. The guessing parameter, in this case, reflects the probability that a state with low "ability" (meaning low adoption of the total number of items) adopts a given item by chance. The discrimination parameter "quantifies how well the item distinguishes between subjects with low/high standing in the latent scale" (Rizopoulos 2006). Finally, the difficulty parameter measures the difficulty of each particular item. I estimated four different Rasch Models: (1) the discriminate is constrained to equal 1; (2) an unconstrained discriminate;

(3) adds a discriminant for each item; and (4) includes a guessing parameter in the unconstrained Rasch Model. Using likelihood ratio tests, I find that option (2) provides the best fit relative to the other three options (see Table S3), thus I use this method for estimating comprehensiveness.[1] Finally, I use multiple imputation for empty cells when generating the factor scores of comprehensiveness due to the small sample properties of the data (Rizopoulos and Moustaki 2008).

Table S3
Comparison of model fit for each estimated Rasch model

Model	AIC	BIC	Log Likelihood	Difference in Log Likelihood
Constrained Discriminant	694	724	-331.81	22.41**
Unconstrained Discriminant	673	705	-320.61	
Discriminant for Each Item	681	742	-308.36	22.48*
Unconstrained Discriminant with Guessing Parameter	699	762	-316.32	6.58

* indicates significance difference from the model with an unconstrained discriminant at * p <0.10 (two-tailed); **p<0.05

References

R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rizopoulos, Dimitris. 2006. "Ltm: An R Package for Latent Variable Modeling and Item Response Analysis." *Journal of Statistical Software* 17: 1-25.

Rizopoulos, Dimitris, and Irini Moustaki. 2008. "Generalized Latent Variable Models with Non-Linear Effects." *British Journal of Mathematical and Statistical Psychology* 61: 415-38.

[1] While the difference in fit with the model including a guessing parameter (4) is not statistically significant, the guessing parameters for only two policies are themselves statistically significant (district policy and legal remedies). Code for generating all four models will be included in the reproduction materials.