Simulation results to accompany:

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (in press). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods*.

Here we report the results of three targeted simulation studies comparing the use of LMS to observed cluster means (termed UMM for "unconflated multilevel model" in Preacher et al., 2011) rather than latent cluster means.

Simulation 1

The first simulation uses the following conditions:

No. of clusters:	50, 100, 200
Cluster size:	5, 10, 20
Hypothesis:	B2 (interaction of L2 variable with latent cluster mean of a L1 variable)
ICC:	ICC = .5 for x_{ij}
Effect size:	Interaction effect = .2
Simulation reps:	500 per cell

For each of the 18 cells of the design, we examined bias in the estimate (mean estimated interaction effect vs. the true value of .2) and bias in the estimated standard error (mean estimated *SE* vs. the empirical standard deviation [ESD] of the estimate across reps). All 500 reps converged for all cells of the design. The following tables summarize the results:

PRB for interaction		Estimate		PRB	
J	nJ	LMS	UMM	LMS	UMM
50	5	0.1972	0.1615	-1.4	-19.3
100	5	0.1994	0.1638	-0.3	-18.1
200	5	0.1939	0.1584	-3.1	-20.8
50	10	0.2038	0.1815	1.9	-9.3
100	10	0.1894	0.1696	-5.3	-15.2
200	10	0.2003	0.1787	0.1	-10.7
50	20	0.2016	0.1884	0.8	-5.8
100	20	0.2049	0.1911	2.4	-4.5
200	20	0.2027	0.1886	1.3	-5.7

Note. PRB = percent relative bias; LMS = latent moderated structural equations; UMM = unconflated measured manifest means; J = number of clusters; n_J = cluster size.

CI coverage			
J	nJ	LMS	UMM
50	5	0.888	0.872
100	5	0.924	0.908
200	5	0.946	0.908
50	10	0.910	0.898
100	10	0.904	0.880
200	10	0.930	0.926
50	20	0.878	0.870
100	20	0.938	0.940
200	20	0.926	0.928

ESD for interaction			
J	nJ	LMS	UMM
50	5	0.2358	0.1902
100	5	0.1553	0.1289
200	5	0.1018	0.0832
50	10	0.2131	0.1893
100	10	0.1410	0.1264
200	10	0.0998	0.0892
50	20	0.2193	0.2052
100	20	0.1271	0.1180
200	20	0.0951	0.0882

Note. ESD = empirical standard deviation.

SE avg. for interaction				PRB for SE	
J	nJ	LMS	UMM	LMS	UMM
50	5	0.2055	0.1633	-12.8	-14.1
100	5	0.1444	0.1162	-7.0	-9.9
200	5	0.1010	0.0818	-0.8	-1.7
50	10	0.1871	0.1635	-12.2	-13.6
100	10	0.1297	0.1145	-8.0	-9.4
200	10	0.0940	0.0830	-5.8	-7.0
50	20	0.1795	0.1656	-18.1	-19.3
100	20	0.1251	0.1158	-1.6	-1.9
200	20	0.0898	0.0831	-5.6	-5.8

MSE			
J	nJ	LMS	UMM
50	5	0.0555	0.0376
100	5	0.0241	0.0179
200	5	0.0104	0.0086
50	10	0.0453	0.0361
100	10	0.0200	0.0169
200	10	0.0099	0.0084
50	20	0.0480	0.0421
100	20	0.0161	0.0140
200	20	0.0090	0.0079

Note. MSE = mean squared error.

To summarize the results, LMS is superior in terms of minimizing bias and achieving more accurate CI coverage, whereas UMM is superior in terms of efficiency. UMM also appears superior in terms of MSE (the combination of bias and sampling variance), but this is largely due to UMM's greater underestimation of its ESD (see "PRB for SE" columns).

UMM approaches LMS's performance in terms of bias and CI coverage more closely as the sample size increases, as would be expected from prior research comparing these methods in other contexts (e.g., Preacher et al., 2011). However, under the limited conditions examined, bias never reached acceptable levels for UMM. On the basis of our results, MSEM with our proposed LMS method is arguably superior to the prevailing popular method of using observed cluster means. Note that our simulations used a predictor ICC of .5, which is quite high. The relative performance of UMM will suffer more as ICC decreases to levels more commonly encountered in practice (Preacher et al., 2011). These results mirror similar simulation results presented by Preacher et al. (2011) and Lüdtke et al. (2008), and serve to support our contention that our recommended LMS approach may be useful for researchers in practice.

Simulation 2

To investigate the costs of unbalanced cluster sizes, we repeated our simulation with a modification. We used the same conditions as before, including the same total sample sizes, but arrived at these total sample sizes using unbalanced clusters. For instance, rather than 50 clusters of size 5, we used 20 clusters of size 2, 18 clusters of size 5, and 10 clusters of size 12 (other conditions used this same cluster ratio of 20:18:10 to maintain consistency across cells). All 500 reps converged in each condition.

Cluster sizes for unbalanced conditions							
				2	**5**	**12**	**Tot**
50	5	250	20	18	10	250
100	5	500	40	36	20	500
200	5	1000	80	72	40	1000
			4	**10**	**24**	**Tot**
50	10	500	20	18	10	500
100	10	1000	40	36	20	1000
200	10	2000	80	72	40	2000
			8	**20**	**48**	**Tot**
50	20	1000	20	18	10	1000
100	20	2000	40	36	20	2000
200	20	4000	80	72	40	4000

PRB for interaction		Estimate		PRB	
J	nJ	LMS	UMM	LMS	UMM
50	5	0.1852	0.1493	-7.4	-25.4
100	5	0.1927	0.1478	-3.7	-26.1
200	5	0.1985	0.1526	-0.8	-23.7
50	10	0.2031	0.1735	1.6	-13.3
100	10	0.1919	0.1653	-4.1	-17.4
200	10	0.2037	0.1752	1.8	-12.4
50	20	0.2113	0.1928	5.6	-3.6
100	20	0.2195	0.2006	9.7	0.3
200	20	0.1992	0.1816	-0.4	-9.2

CI coverage			
J	nJ	LMS	UMM
50	5	0.876	0.876
100	5	0.904	0.886
200	5	0.928	0.890
50	10	0.890	0.882
100	10	0.932	0.910
200	10	0.944	0.934
50	20	0.888	0.878
100	20	0.904	0.902
200	20	0.928	0.930

ESD for interaction			
J	nJ	LMS	UMM
50	5	0.2852	0.2085
100	5	0.1718	0.1328
200	5	0.1173	0.0890
50	10	0.2472	0.2101
100	10	0.1455	0.1253
200	10	0.0999	0.0859
50	20	0.2219	0.2020
100	20	0.1474	0.1351
200	20	0.0992	0.0905

Note. ESD = empirical standard deviation.

SE avg. for interaction				PRB for SE	
J	nJ	LMS	UMM	LMS	UMM
50	5	0.2306	0.1698	-19.1	-18.6
100	5	0.1566	0.1189	-8.8	-10.5
200	5	0.1106	0.0845	-5.7	-5.1
50	10	0.1987	0.1668	-19.6	-20.6
100	10	0.1395	0.1185	-4.1	-5.4
200	10	0.0992	0.0845	-0.7	-1.6
50	20	0.1865	0.1671	-16.0	-17.3
100	20	0.1310	0.1182	-11.1	-12.5
200	20	0.0942	0.0853	-5.0	-5.7

MSE			
J	nJ	LMS	UMM
50	5	0.0814	0.0459
100	5	0.0295	0.0202
200	5	0.0137	0.0102
50	10	0.0610	0.0448
100	10	0.0212	0.0169
200	10	0.0100	0.0080
50	20	0.0493	0.0408
100	20	0.0221	0.0182
200	20	0.0098	0.0085

Note. MSE = mean squared error.

CI coverage is about the same as with balanced clusters. Bias, ESD, *SE*, and MSE were larger with unbalanced clusters for nearly all conditions. However, the relative performance of UMM and LMS remained the same: LMS showed less bias and more accurate coverage, while UMM was more efficient.

Simulation 3

In the third simulation, we compared LMS to using cluster means (UMM), but we changed the ICC of x_{ij} to .1 rather than .5. Furthermore, the current simulation was equated to the first simulation in all respects other than the ICC of x_{ij} by reparameterizing the model to contain standardized effects. This allowed us to use all the same parameter values across simulations, changing only the proportions of 'within' and 'between' variance in x_{ij}.

Number of clusters:	50, 100, 200
Cluster size:	5, 10, 20 (balanced)
Hypothesis:	B2 (intxn of L2 variable with latent cluster mean of L1 variable)
ICCs:	ICC = .1 for x_{ij}
Effect size:	Interaction effect = .4472 (after rescaling)
Simulation reps:	500 per cell

We added another condition (J = 200, n_j = 80) to examine what would happen under conditions of very large clusters. For each of the 20 cells of the design, we examined bias in the interaction effect and in the estimated standard error (mean estimated *SE* vs. the empirical standard deviation [ESD] of the estimate across reps). All runs converged for both LMS and UMM, with the exception of the J = 50, n_j = 5 cell for LMS (494 out of 500 converged). The following tables summarize the results. For reference, the first table in each pair reports the results of the previous simulation, but note both simulations involved unstandardized effects that are necessarily different (.200 for the original simulation, and .447 for the new one).

PRB for interaction		Estimate		PRB	
J	nJ	LMS	UMM	LMS	UMM
50	5	0.1972	0.1615	-1.4	-19.3
100	5	0.1994	0.1638	-0.3	-18.1
200	5	0.1939	0.1584	-3.1	-20.8
50	10	0.2038	0.1815	1.9	-9.3
100	10	0.1894	0.1696	-5.3	-15.2
200	10	0.2003	0.1787	0.1	-10.7
50	20	0.2016	0.1884	0.8	-5.8
100	20	0.1236	0.1911	-11.7	-4.5
200	20	0.2027	0.1886	1.3	-5.7

PRB for interaction		Estimate		PRB	
J	nJ	LMS	UMM	LMS	UMM
50	5	0.4250	0.1539	-5.0	-65.6
100	5	0.4276	0.1590	-4.4	-64.4
200	5	0.4242	0.1550	-5.1	-65.3
50	10	0.4693	0.2361	4.9	-47.2
100	10	0.4466	0.2206	-0.1	-50.7
200	10	0.4696	0.2337	5.0	-47.7
50	20	0.4342	0.3126	-2.9	-30.1
100	20	0.4312	0.3168	-3.6	-29.2
200	20	0.4305	0.3069	-3.7	-31.4
200	80	0.4477	0.4077	0.1	-8.8

Note. PRB = percent relative bias; LMS = latent moderated structural equations; UMM = unconflated measured manifest means; *J* = number of clusters; n_J = cluster size.

CI coverage			
J	nJ	LMS	UMM
50	5	0.888	0.872
100	5	0.924	0.908
200	5	0.946	0.908
50	10	0.910	0.898
100	10	0.904	0.880
200	10	0.930	0.926
50	20	0.878	0.870
100	20	0.938	0.940
200	20	0.926	0.928

CI coverage			
J	nJ	LMS	UMM
50	5	0.872	0.704
100	5	0.906	0.576
200	5	0.932	0.352
50	10	0.886	0.834
100	10	0.932	0.730
200	10	0.944	0.670
50	20	0.878	0.862
100	20	0.930	0.894
200	20	0.944	0.838
200	80	0.926	0.904

ESD for interaction			
J	nJ	LMS	UMM
50	5	0.2358	0.1902
100	5	0.1553	0.1289
200	5	0.1018	0.0832
50	10	0.2131	0.1893
100	10	0.1410	0.1264
200	10	0.0998	0.0892
50	20	0.2193	0.2052
100	20	0.1271	0.1180
200	20	0.0951	0.0882

ESD for interaction			
J	nJ	LMS	UMM
50	5	1.0917	0.2798
100	5	0.7161	0.1865
200	5	0.3917	0.1260
50	10	0.7286	0.3219
100	10	0.4258	0.2181
200	10	0.2696	0.1502
50	20	0.5675	0.3792
100	20	0.3465	0.2242
200	20	0.2347	0.1673
200	80	0.2143	0.1941

Note. ESD = empirical standard deviation.

SE avg. for interaction				PRB for SE	
J	nJ	LMS	UMM	LMS	UMM
50	5	0.2055	0.1633	-12.8	-14.1
100	5	0.1444	0.1162	-7.0	-9.9
200	5	0.1010	0.0818	-0.8	-1.7
50	10	0.1871	0.1635	-12.2	-13.6
100	10	0.1297	0.1145	-8.0	-9.4
200	10	0.0940	0.0830	-5.8	-7.0
50	20	0.1795	0.1656	-18.1	-19.3
100	20	0.1251	0.1158	-1.6	-1.9
200	20	0.0898	0.0831	-5.6	-5.8

SE avg. for interaction				PRB for SE	
J	nJ	LMS	UMM	LMS	UMM
50	5	0.9695	0.2410	-11.2	-13.9
100	5	0.6712	0.1730	-6.3	-7.2
200	5	0.3919	0.1228	0.1	-2.5
50	10	0.6224	0.2824	-14.6	-12.3
100	10	0.4099	0.1973	-3.7	-9.5
200	10	0.2737	0.1427	1.5	-5.0
50	20	0.4855	0.3180	-14.4	-16.1
100	20	0.3285	0.2218	-5.2	-1.1
200	20	0.2308	0.1594	-1.7	-4.7
200	80	0.1994	0.1731	-7.0	-10.8

MSE				MSE			
J	nJ	LMS	UMM	J	nJ	LMS	UMM
50	5	0.0555	0.0376	50	5	1.1898	0.1641
100	5	0.0241	0.0179	100	5	0.5122	0.1178
200	5	0.0104	0.0086	200	5	0.1537	0.1012
50	10	0.0453	0.0361	50	10	0.5303	0.1480
100	10	0.0200	0.0169	100	10	0.1809	0.0988
200	10	0.0099	0.0084	200	10	0.0731	0.0681
50	20	0.0480	0.0421	50	20	0.3216	0.1616
100	20	0.0161	0.0140	100	20	0.1201	0.0672
200	20	0.0090	0.0079	200	20	0.0552	0.0476
				200	80	0.0458	0.0392

Note. MSE = mean squared error.

To summarize the results, as in the previous simulation, LMS is superior in terms of minimizing bias and achieving more accurate CI coverage. Across all conditions, the bias associated with UMM is unacceptably large. CI coverage is also unacceptably low for UMM in most conditions. Except for large cluster size conditions, UMM's CI coverage gets worse as the number of clusters increases because the CIs become narrower around the more heavily biased point estimates.

UMM appears superior in terms of efficiency—although, as we note below, this does not translate into higher statistical power. This is partly due to the overall smaller effects, around which there is likely to be less uncertainty, and partly due to the fact that UMM's *SE*s underestimate the ESDs more than does LMS in most conditions. UMM also appears superior in terms of MSE (the combination of bias and sampling variance); this is driven by the much smaller standard errors, which are able to compensate for the large bias.

In summary, there are stark differences between LMS and UMM in the limited conditions examined here. These differences can be seen as an example of the bias-variance trade-off. LMS minimizes bias at the cost of efficiency, which is reduced. UMM sacrifices unbiasedness in return for greater efficiency. In our view, the bias associated with UMM renders it unusable, whereas the inefficiency associated with LMS does not render it unusable (i.e., we consider bias to be more of a problem than uncertainty). Interestingly, the superior efficiency of UMM does not translate to markedly, or even uniformly, higher power:

Power			
J	nJ	LMS	UMM
50	5	0.134	0.144
100	5	0.174	0.192
200	5	0.262	0.256
50	10	0.174	0.192
100	10	0.238	0.238
200	10	0.428	0.406
50	20	0.230	0.238
100	20	0.314	0.324
200	20	0.476	0.498
200	80	0.609	0.638

This is a telling result. If a researcher is trying to test a null hypothesis about the parameter, it may not matter much which method is used. If the researcher is trying to estimate the parameter, LMS gives much lower bias and more accurate CI coverage. This is an instance where lower MSE can be quite misleading about the quality of a method.