Overview

The contents of this document contain the Online Supplementary materials to "Interrater reliability for Multilevel Data: A Generalizability Theory Approach". Additional materials, such as the Stan models, the R syntax from the simulations and the applied example, and the data used for the applied example are available through the Open Science Framework <URL masked for review>.

S1. Population Values ICCs Simulation 1

Table 1

Population Values of Subject-Level ICCs across Conditions

	$\sigma_c^2 = 0.16$		$\sigma_c^2 = 0.50$		$\sigma_c^2 = 0.16$		$\sigma_c^2 = 0.50$	
k	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$
	$\text{ICC}_s(A, 1)$				$\text{ICC}_s(A, k)$			
2	0.60	0.50	0.60	0.50	0.75	0.67	0.75	0.67
5	0.60	0.50	0.60	0.50	0.88	0.83	0.88	0.83
10	0.60	0.50	0.60	0.50	0.94	0.91	0.94	0.91
	$\text{ICC}_s(C, 1)$				$\text{ICC}_s(C, k)$			
2	0.67	0.67	0.67	0.67	0.80	0.80	0.80	0.80
5	0.67	0.67	0.67	0.67	0.91	0.91	0.91	0.91
10	0.67	0.67	0.67	0.67	0.95	0.95	0.95	0.95

ICC = Intraclass correlation; k = Number of raters; σ_c^2 = Variance of cluster effects; σ_r^2 = Variance of rater effects.

Table 2

Population Values of Cluster-Level ICCs across Conditions

	$\sigma_c^2 = 0.16$		$\sigma_c^2 = 0.50$		$\sigma_c^2 = 0.16$		$\sigma_c^2 = 0.50$	
k	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$
	$\mathrm{ICC}_c(A, 1)$				$\mathrm{ICC}_c(A, k)$			
2	0.33	0.20	0.61	0.43	0.50	0.33	0.76	0.60
5	0.33	0.20	0.61	0.43	0.71	0.55	0.89	0.79
10	0.33	0.20	0.61	0.43	0.83	0.71	0.94	0.88
	$\mathrm{ICC}_c(C, 1)$				$\mathrm{ICC}_c(C, k)$			
2	0.50	0.50	0.76	0.76	0.67	0.67	0.86	0.86
5	0.50	0.50	0.76	0.76	0.83	0.83	0.94	0.94
10	0.50	0.50	0.76	0.76	0.91	0.91	0.97	0.97

ICC = Intraclass correlation; k = Number of raters; σ_c^2 = Variance of cluster effects; σ_r^2 = Variance of rater effects.

Table 3

Relative Difference in Population Values of Conflated ICCs compared to Multilevel ICCs across Conditions

	σ_c^2= 0.16		σ_c^2= 0.50		σ_c^2= 0.16		σ_c^2= 0.50	
k	σ_r^2= 0.16	σ_r^2= 0.50	σ_r^2= 0.16	σ_r^2= 0.50	σ_r^2= 0.16	σ_r^2= 0.50	σ_r^2= 0.16	σ_r^2= 0.50
				Subject-Level				height
	$\text{ICC}_s(A,1)$				$\text{ICC}_s(A,k)$			
2	2.51	4.88	9.27	15.35	1.00	1.00	5.60	5.60
5	2.51	4.88	9.27	15.35	1.00	1.00	5.60	5.60
10	2.51	4.88	9.27	15.35	1.00	1.00	5.60	5.60
		$\text{ICC}_s(C,1)$				$\text{ICC}_s(C,k)$		
2	1.45	3.05	5.22	9.25	0.55	0.55	3.03	3.03
5	0.64	1.44	2.26	4.22	0.24	0.24	1.28	1.28
10	0.33	0.76	1.16	2.21	0.12	0.12	0.65	0.65
				Cluster-Level				height
	$\text{ICC}_c(A,1)$				$\text{ICC}_c(A,k)$			
2	57.03	156.80	6.26	35.21	-6.27	-6.27	-11.20	-11.20
5	57.03	156.80	6.26	35.21	-6.27	-6.27	-11.20	-11.20
10	57.03	156.80	6.26	35.21	-6.27	-6.27	-11.20	-11.20
		$\text{ICC}_c(C,1)$				$\text{ICC}_c(C,k)$		
2	33.02	98.04	3.53	21.22	-3.47	-3.47	-6.07	-6.07
5	14.59	46.16	1.53	9.68	-1.48	-1.48	-2.56	-2.56
10	7.56	24.52	0.79	5.08	-0.76	-0.76	-1.30	-1.30

ICC = Intraclass correlation; k = Number of raters; σ_c^2 = Variance of cluster effects; σ_r^2 = Variance of rater effects. Let ICC_{conf} denote the conflated ICC, and let ICC_{ML} the multilevel ICC. Relative difference is computed as $\frac{\text{ICC}_{conf}-\text{ICC}_{ML}}{\text{ICC}_{ML}} * 100$.

S2. Results Simulation 1

Table 4
Relative MAP Bias of Subject-Level ICCs across Conditions

			σ_c^2= 0.16		σ_c^2= 0.50		σ_c^2= 0.16		σ_c^2= 0.50	
N_c	N_s	K	σ_r^2= 0.16	σ_r^2= 0.50	σ_r^2= 0.16	σ_r^2= 0.50	σ_r^2= 0.16	σ_r^2= 0.50	σ_r^2= 0.16	σ_r^2= 0.50
			$\text{ICC}_s(A,1)$				$\text{ICC}_s(A,k)$			
20	10	2	-0.03	0.06	-0.02	0.06	-0.01	0.04	-0.01	0.04
		5	-0.01	0.03	0.00	0.02	0.00	0.01	0.00	0.01
		10	-0.01	0.01	-0.01	0.01	0.00	0.00	0.00	0.00
	30	2	0.00	0.10	0.02	0.10	0.00	0.07	0.01	0.07
		5	0.00	0.03	0.01	0.03	0.00	0.01	0.00	0.01
		10	0.00	0.02	0.00	0.02	0.00	0.00	0.00	0.00
40	10	2	-0.02	0.06	-0.02	0.06	-0.01	0.05	-0.01	0.05
		5	0.00	0.03	0.00	0.02	0.00	0.01	0.00	0.01
		10	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00
	30	2	0.00	0.09	0.02	0.09	0.00	0.06	0.01	0.06
		5	0.00	0.03	0.01	0.03	0.00	0.01	0.00	0.01
		10	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00
			$\text{ICC}_s(C,1)$				$\text{ICC}_s(C,k)$			
20	10	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	30	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40	10	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	30	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

ICC = Intraclass correlation; N_c = Number of clusters; N_s = Number of subjects per cluster; K = Number of raters; σ_c^2 = Variance of cluster effects; σ_r^2 = Variance of rater effects. Relative bias was computed as $\frac{\bar{\theta}-\theta}{\theta}$, where $\bar{\theta}$ denotes the average MAP estimate of a parameter (or derived ICC) across replications in a condition, and θ denotes the population parameter in that condition.

Table 5

Relative MAP Bias of Cluster-Level ICCs across Conditions

			$\sigma_c^2 = 0.16$		$\sigma_c^2 = 0.50$		$\sigma_c^2 = 0.16$		$\sigma_c^2 = 0.50$	
N_c	N_s	K	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$
			$\text{ICC}_c(A,1)$				$\text{ICC}_c(A,k)$			
20	10	2	-0.54	-0.41	-0.22	-0.11	-0.42	-0.31	-0.13	-0.05
		5	-0.25	-0.23	-0.06	-0.05	-0.11	-0.13	-0.01	0.00
		10	-0.14	-0.15	-0.04	-0.04	-0.02	-0.04	0.00	0.00
	30	2	-0.53	-0.38	-0.14	-0.04	-0.4	-0.29	-0.06	0.00
		5	-0.15	-0.12	-0.04	-0.02	-0.05	-0.04	-0.01	0.00
		10	-0.08	-0.07	-0.03	-0.02	-0.01	-0.01	0.00	0.00
40	10	2	-0.48	-0.33	-0.13	-0.02	-0.36	-0.25	-0.06	0.01
		5	-0.18	-0.16	-0.04	-0.01	-0.06	-0.06	-0.01	0.01
		10	-0.09	-0.10	-0.02	-0.02	-0.01	-0.02	0.00	0.00
	30	2	-0.41	-0.26	-0.05	0.04	-0.27	-0.15	-0.01	0.06
		5	-0.10	-0.05	-0.02	0.01	-0.03	0.00	0.00	0.02
		10	-0.06	-0.04	-0.01	-0.01	-0.01	0.00	0.00	0.00
			$\text{ICC}_c(C,1)$				$\text{ICC}_c(C,k)$			
20	10	2	-0.13	-0.13	-0.02	-0.02	-0.08	-0.08	-0.01	-0.01
		5	-0.09	-0.09	0.00	-0.01	-0.02	-0.02	0.00	0.00
		10	-0.05	-0.05	0.00	0.00	0.00	0.00	0.00	0.00
	30	2	-0.16	-0.16	-0.01	-0.02	-0.10	-0.11	0.00	-0.01
		5	-0.02	-0.02	0.00	0.00	0.00	0.00	0.00	0.00
		10	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00
40	10	2	-0.09	-0.08	0.00	0.00	-0.05	-0.05	0.00	0.00
		5	-0.04	-0.04	0.00	0.00	-0.01	-0.01	0.00	0.00
		10	-0.01	-0.02	0.00	0.00	0.00	0.00	0.00	0.00
	30	2	-0.03	-0.03	0.01	0.00	-0.01	-0.02	0.01	0.00
		5	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00
		10	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00

ICC = Intraclass correlation; N_c = Number of clusters; N_s = Number of subjects per cluster; K = Number of raters; σ_c^2 = Variance of cluster effects; σ_r^2 = Variance of rater effects. Relative bias was computed as $\frac{\bar{\theta}-\theta}{\theta}$, where $\bar{\theta}$ denotes the average MAP estimate of a parameter (or derived ICC) across replications in a condition, and θ denotes the population parameter in that condition.

Table 6

Percentage of 95% BCI Coverage of Subject-Level ICCs across Conditions

			$\sigma_c^2 = 0.16$		$\sigma_c^2 = 0.50$		$\sigma_c^2 = 0.16$		$\sigma_c^2 = 0.50$	
N_c	$N_{s:c}$	K	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$
			$\text{ICC}_s(A,1)$				$\text{ICC}_s(A,k)$			
20	10	2	95	98*	95	98*	95	98*	95	98*
		5	95	96	95	96	95	96	95	96
		10	95	95	95	95	95	95	95	95
	30	2	97*	98*	97*	98*	97*	98*	97*	98*
		5	94*	95	94	95	94*	95	94	95
		10	94	96	95	95	94	96	95	95
40	10	2	96*	98*	96	98*	96*	98*	96	98*
		5	95	97*	95	97*	95	97*	95	97*
		10	95	95	95	95	95	95	95	95
	30	2	97*	99*	98*	99*	97*	99*	98*	99*
		5	95	97*	95	97*	95	97*	95	97*
		10	95	96	96	96	95	96	96	96
			$\text{ICC}_s(C,1)$				$\text{ICC}_s(C,k)$			
20	10	2	94	95	94	94	94	95	94	94
		5	95	95	95	95	95	95	95	95
		10	95	95	95	95	95	95	95	95
	30	2	97*	96	97*	96*	97*	96	97*	96*
		5	96	95	96	95	96	95	96	95
		10	95	94	94	94	95	94	94	94
40	10	2	96	96	96	96	96	96	96	96
		5	94	94	95	94	94	94	95	94
		10	94	94	94	94	94	94	94	94
	30	2	96	96	96	96	96	96	96	96
		5	94	95	95	95	94	95	95	95
		10	95	95	95	95	95	95	95	95

ICC = Intraclass correlation; N_c = Number of clusters; N_s = Number of subjects per cluster; K = Number of raters; σ_c^2 = Variance of cluster effects; σ_s^2 = Variance of subject effects; * Outside 95% Agresti-Coull confidence interval.

Table 7

Percentage of 95% BCI Coverage of Cluster-Level ICCs across Conditions

			σ^2_c= 0.16		σ^2_c= 0.50		σ^2_c= 0.16		σ^2_c= 0.50	
N_c	$N_{s:c}$	K	σ^2_r= 0.16	σ^2_r= 0.50	σ^2_r= 0.16	σ^2_r= 0.50	σ^2_r= 0.16	σ^2_r= 0.50	σ^2_r= 0.16	σ^2_r= 0.50
			$\text{ICC}_c(A,1)$				$\text{ICC}_c(A,k)$			
20	10	2	96	98*	94	97*	96	98*	94	97*
		5	95	96	96	96	95	96	96	96
		10	96	96	96*	96	96	96	96*	96
	30	2	95	98*	94	97*	95	98*	94	97*
		5	95	96	95	96	95	96	95	96
		10	95	95	95	95	95	95	95	95
40	10	2	94	96	95	97*	94	96	95	97*
		5	94	95	95	96*	94	95	95	96*
		10	94	94	95	96	94	94	95	96
	30	2	96	97*	96	98*	96	97*	96	98*
		5	96	97*	96*	97*	96	97*	96*	97*
		10	95	95	95	95	95	95	95	95
			$\text{ICC}_c(C,1)$				$\text{ICC}_c(C,k)$			
20	10	2	97*	96*	94*	94*	97*	96*	94*	94*
		5	95	95	95	95	95	95	95	95
		10	96	96	96	96*	96	96	96	96*
	30	2	96	96	94	95	96	96	94	95
		5	95	95	96	95	95	95	96	95
		10	95	95	95	96	95	95	95	96
40	10	2	95	95	94	94	95	95	94	94
		5	95	94	96	95	95	94	96	95
		10	95	94	95	95	95	94	95	95
	30	2	94	95	94	95	94	95	94	95
		5	95	94	96	96	95	94	96	96
		10	95	95	95	95	95	95	95	95

ICC = Intraclass correlation; N_c = Number of clusters; N_s = Number of subjects per cluster; K = Number of raters; σ^2_c = Variance of cluster effects; σ^2_s = Variance of subject effects; * Outside 95% Agresti-Coull confidence interval.

Table 8

Relative Efficiency of Subject-Level ICCs across Conditions

			$\sigma_c^2 = 0.16$		$\sigma_c^2 = 0.50$		$\sigma_c^2 = 0.16$		$\sigma_c^2 = 0.50$	
N_c	N_s	K	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$	$\sigma_r^2 = 0.16$	$\sigma_r^2 = 0.50$
			$\text{ICC}_s(A,1)$				$\text{ICC}_s(A,k)$			
20	10	2	1.98	1.51	1.96	1.51	2.15	1.62	2.13	1.62
		5	1.37	1.22	1.37	1.22	1.58	1.40	1.58	1.39
		10	1.13	1.08	1.12	1.08	1.21	1.20	1.20	1.19
	30	2	2.22	1.61	2.22	1.62	2.39	1.72	2.39	1.73
		5	1.37	1.18	1.38	1.18	1.59	1.37	1.59	1.37
		10	1.15	1.07	1.14	1.07	1.24	1.18	1.24	1.18
40	10	2	2.15	1.59	2.13	1.59	2.33	1.71	2.31	1.71
		5	1.44	1.25	1.44	1.25	1.66	1.43	1.66	1.44
		10	1.21	1.14	1.20	1.14	1.30	1.26	1.29	1.26
	30	2	2.24	1.66	2.25	1.66	2.44	1.78	2.45	1.79
		5	1.46	1.24	1.46	1.24	1.69	1.44	1.70	1.43
		10	1.24	1.15	1.24	1.15	1.35	1.28	1.35	1.28
			$\text{ICC}_s(C,1)$				$\text{ICC}_s(C,k)$			
20	10	2	0.98	0.99	0.98	0.98	0.98	0.99	0.98	0.98
		5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		10	1.01	1.02	1.02	1.02	1.01	1.01	1.01	1.01
	30	2	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
		5	1.02	1.02	1.02	1.02	1.01	1.01	1.02	1.02
		10	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
40	10	2	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03
		5	0.99	0.99	0.98	0.99	0.99	0.99	0.98	0.99
		10	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.97
	30	2	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
		5	1.03	1.03	1.03	1.02	1.03	1.03	1.03	1.02
		10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

ICC = Intraclass correlation; N_c = Number of clusters; N_s = Number of subjects per cluster; K = Number of raters; σ_c^2 = Variance of cluster effects; σ_r^2 = Variance of rater effects. Relative efficiency was computed as the ratio of the average posterior *SD* relative to the *SD* of the posterior means. Preferably, this ratio equals 1.

Table 9

Relative Efficiency of Cluster-Level ICCs across Conditions

			σ_c^2= 0.16		σ_c^2= 0.50		σ_c^2= 0.16		σ_c^2= 0.50	
N_c	N_s	K	σ_r^2= 0.16	σ_r^2= 0.50	σ_r^2= 0.16	σ_r^2= 0.50	σ_r^2= 0.16	σ_r^2= 0.50	σ_r^2= 0.16	σ_r^2= 0.50
			$\text{ICC}_c(A,1)$				$\text{ICC}_c(A,k)$			
20	10	2	1.38	1.26	1.40	1.24	1.41	1.29	1.42	1.26
		5	1.06	1.02	1.11	1.06	1.09	1.06	1.16	1.12
		10	1.04	1.04	1.04	1.04	1.00	1.01	0.97	1.00
	30	2	1.40	1.26	1.53	1.32	1.43	1.29	1.58	1.37
		5	1.10	1.04	1.15	1.07	1.09	1.06	1.22	1.15
		10	1.03	1.01	1.03	1.02	0.96	0.99	1.01	1.03
40	10	2	1.33	1.18	1.57	1.31	1.37	1.22	1.68	1.39
		5	1.09	1.03	1.20	1.10	1.11	1.07	1.34	1.22
		10	1.01	0.99	1.05	1.04	0.95	0.97	1.06	1.07
	30	2	1.47	1.27	1.78	1.43	1.52	1.32	1.91	1.53
		5	1.13	1.05	1.24	1.11	1.21	1.12	1.42	1.26
		10	1.05	1.01	1.10	1.06	1.06	1.06	1.15	1.13
			$\text{ICC}_c(C,1)$				$\text{ICC}_c(C,k)$			
20	10	2	1.14	1.14	0.96	0.97	1.21	1.21	0.97	0.97
		5	1.00	0.99	0.92	0.93	1.05	1.04	0.88	0.88
		10	0.99	0.99	0.94	0.93	1.00	0.99	0.81	0.77
	30	2	1.10	1.08	1.02	1.02	1.14	1.13	1.03	1.03
		5	0.97	0.97	0.97	0.96	0.90	0.90	0.92	0.92
		10	0.99	0.99	0.96	0.97	0.85	0.84	0.90	0.91
40	10	2	1.03	1.03	0.97	0.98	1.06	1.06	0.96	0.97
		5	0.97	0.96	0.96	0.96	0.93	0.92	0.92	0.91
		10	0.95	0.94	0.97	0.98	0.87	0.84	0.91	0.92
	30	2	1.01	1.00	1.00	1.00	1.02	1.01	1.00	1.00
		5	0.99	0.99	0.99	1.00	0.94	0.94	0.97	0.98
		10	1.01	1.00	1.00	1.00	0.93	0.93	0.97	0.97

ICC = Intraclass correlation; N_c = Number of clusters; N_s = Number of subjects per cluster; K = Number of raters; σ_c^2 = Variance of cluster effects; σ_r^2 = Variance of rater effects. Relative efficiency was computed as the ratio of the average posterior *SD* relative to the *SD* of the posterior means. Preferably, this ratio equals 1.

S3. Results Simulation 2

Table 10

Relative MAP Bias of ICCs Across Conditions

k_s	K	$\text{ICC}_s(A,1)$	$\text{ICC}_s(A,k)$	$\text{ICC}_s(C,1)$	$\text{ICC}_s(C,k)$
2	5	-0.02	-0.01	0.00	0.00
	10	-0.02	-0.01	0.00	0.00
3	5	-0.02	-0.01	0.00	0.00
	10	-0.01	-0.01	0.00	0.00
k_s	K	$\text{ICC}_c(A,1)$	$\text{ICC}_c(A,k)$	$\text{ICC}_c(C,1)$	$\text{ICC}_c(C,k)$
2	5	-0.29	-0.22	-0.08	-0.05
	10	-0.26	-0.20	-0.06	-0.03
3	5	-0.21	-0.12	-0.04	-0.01
	10	-0.21	-0.12	-0.04	-0.01

ICC = Intraclass correlation; k_s = Number of raters per subject; K = Total number of raters (size of the rater pool). Relative bias was computed as $\frac{\bar{\theta}-\theta}{\theta}$, where $\bar{\theta}$ denotes the average MAP estimate of a parameter (or derived ICC) across replications in a condition, and θ denotes the population parameter in that condition.

Table 11

Percentage of 95% BCI Coverage of ICCs Across Conditions

k_s	K	$\text{ICC}_s(A,1)$	$\text{ICC}_s(A,k)$	$\text{ICC}_s(C,1)$	$\text{ICC}_s(C,k)$
2	5	95	95	96	96
	10	95	95	96	96
3	5	95	95	96	96
	10	95	95	96	96
k_s	K	$\text{ICC}_c(A,1)$	$\text{ICC}_c(A,k)$	$\text{ICC}_c(C,1)$	$\text{ICC}_c(C,k)$
2	5	96	96	95	95
	10	95	95	95	95
3	5	95	95	94	94
	10	95	95	96	96

ICC = Intraclass correlation; k_s = Number of raters per subject; K = Total number of raters (size of the rater pool).

Table 12

Relative Efficiency of ICCs Across Conditions

k_s	K	$\text{ICC}_s(A,1)$	$\text{ICC}_s(A,k)$	$\text{ICC}_s(C,1)$	$\text{ICC}_s(C,k)$
2	5	1.29	1.36	1.03	1.04
	10	1.29	1.36	1.03	1.03
3	5	1.37	1.50	1.04	1.03
	10	1.32	1.44	1.00	1.00
k_s	K	$\text{ICC}_c(A,1)$	$\text{ICC}_c(A,k)$	$\text{ICC}_c(C,1)$	$\text{ICC}_c(C,k)$
2	5	1.09	1.11	1.04	1.07
	10	1.10	1.11	1.04	1.06
3	5	1.06	1.05	0.97	0.96
	10	1.07	1.08	1.02	1.02

ICC = Intraclass correlation; k_s = Number of raters per subject; K = Total number of raters (size of the rater pool). Relative efficiency was computed as the ratio of the average posterior *SD* relative to the *SD* of the posterior means. Preferably, this ratio equals 1.

S4. Maximum Likelihood Estimates and additional MCMC Estimates

Application

Table 13

Variance Decomposition as Estimated with MLE

	Conflated	MLM
Student	1.63	1.40
Teacher	–	0.23
Rater	0.26	0.26
Student × Rater	0.48	0.44
Teacher × Rater	–	0.03

MLM = Multilevel Model.

Table 14

IRR coefficients as Estimated with MCMC

	ICC(A,1)		ICC(A,k)		ICC(C,1)		ICC(C,k)	
	MAP	BCI	MAP	BCI	MAP	BCI	MAP	BCI
Conflated (k = 3)	.68	[.49, .75]	.87	[.74, .90]	.77	[.73, .81]	.91	[.89, .93]
Student Level (k = 3)	.66	[.46, .74]	.85	[.72, .89]	.76	[.71, .81]	.91	[.88, .93]
Teacher Level (k = 3)	.39	[.10, .68]	.71	[.24, .86]	.90	[.62, .96]	.96	[.83, .98]
Teacher Level (k = 5)	–	–	.80	[.35, .91]	–	–	.98	[.89, .99]

ICC = Intraclass correlation; MAP = maximum a posteriori; BCI = Bayesian credible interval; k = number of raters.

Table 15

IRR coefficients as Estimated with MLE

	ICC(A,1)	ICC(A,k)	ICC(C,1)	ICC(C,k)
Conflated (k = 3)	.69	.87	.77	.91
Student Level (k = 3)	.67	.86	.76	.90
Teacher Level (k = 3)	.45	.71	.88	.96
Teacher Level (k = 5)	–	.80	–	.97

ICC = Intraclass correlation; k = number of raters.