Online Supplement to

Estimating Trait Emotion Differentiation: How Many Measurement Occasions and Emotion Items are Needed?

Sabrina Ecker[1], Esther Ulitzsch[2], and Tanja Lischetzke[1]

[1] RPTU University Kaiserslautern-Landau

[2] University of Oslo

Contents

A: Overview of the Screened Studies 3

B: Descriptive Statistics of Emotion Variables in the Primary Data Set 9

C: Results Tables and Additional Figures for the Main Analyses in the Primary Data Set 10

D: Sensitivity Analysis Using Fisher's *Z*-transformed Intraclass Correlations in the Primary Data Set 15

E: Sensitivity Analyses Using Ordered Draws of Measurement Occasions in the Primary Data Set 21

F: Sensitivity Analyses Using Different Treatments of Negative Intraclass Correlations in the Primary Data Set 25

G: Sensitivity Analysis Comparing Item Sets with Lowest and Highest Endorsement 35

H: Description of the Replication Data Set 42

J: Results for the Replication Data Set (Replication of Main Analysis) 48

K: Results for Fisher's *Z*-Transformed Intraclass Correlations in the Replication Data Set 59

L: Sensitivity Analyses Drawing Measurement Occasions in Temporal Order in the Replication Data Set 71

M: Sensitivity Analyses Using Different Treatments of Negative Intraclass Correlations in the Replication Data Set 80

N: Sensitivity Analysis Comparing Item Sets with Lowest and Highest Endorsement 100

References 114

A: Overview of the Screened Studies

Table A1

Overview of the Screened Studies to Determine Realistic Numbers of Emotion Items and Measurement Occasions Used to Assess Emotion Differentiation

Source	$N_{\text{Level 2}}$	N_{days}	$N_{\text{occ per day}}$	$N_{\text{occ total}}$	$N_{\text{items negative}}$	$N_{\text{items positive}}$	Response scale
Feldman Barrett et al. (2001)	43	14	1	14	5	4	Likert
Tugade et al. (2004)	130	28	10	280	-	5	Likert
Decker et al. (2008)	138	9	1	9	5	4	Likert
Kashdan et al. (2010)	106	21	6	126	6	-	Likert
Pond et al. (2012) – study 1	199	9	1	9	10	10	Likert
Pond et al. (2012) – study 2	186	21	1	21	5	-	Likert
Pond et al. (2012) – study 3	243	9	1	9	10	10	Likert
Hill and Updegraff (2012)	96	7	6	42	11	10	Likert
Demiralp et al. (2012)	106	7	8	56	7	4	Likert
Zaki et al. (2013)	80	21	5	105	7	-	Likert
Grühn et al. (2013)	109	7	5	35	10	10	Likert
Boden et al. (2013) – study 2	99	15	4	60	6	6	Likert

Table A1 (continued)

Source	$N_{\text{Level 2}}$	N_{days}	$N_{\text{occ per day}}$	$N_{\text{occ total}}$	$N_{\text{items negative}}$	$N_{\text{items positive}}$	Response scale
Kashdan and Farmer (2014) – random-prompt index	85	14	5	70	4	4	Likert
Kashdan and Farmer (2014) – social-interaction index	85	14	-	10[a]	4	4	Likert
Kashdan and Farmer (2014) – end-of-day index	85	14	1	14	6	6	Likert
Kimhy et al. (2014)	99	2	10	20	3	-	slider
O'Toole et al. (2014)	164	11	1	11	5	5	Likert
Kashdan et al. (2014)	25	21	1	21	10	-	Likert
Bresin (2014)	118	14	1	14	10	10	Likert
Selby et al. (2014)	118	14	9.09[b]	127.26[b]	15	9	Likert
Vandercammen et al. (2014) – study 1	72	10	1	10	6	6	Likert
Vandercammen et al. (2014) – study 2	34	5	4	20	6	6	Likert
Dixon-Gordon et al. (2014)	84	1	8	8	11	10	Likert
Emery et al. (2014)	102	28	8	224	9	5	Likert
Erbas et al. (2014) – study 1	50	7	10	70	4	-	Likert
Tomko et al. (2015)	105	28	6	168	17	-	Likert

Table A1 (continued)

Source	$N_{\text{Level 2}}$	N_{days}	$N_{\text{occ per day}}$	$N_{\text{occ total}}$	$N_{\text{items negative}}$	$N_{\text{items positive}}$	Response scale
Erbas et al. (2015) – study 1	79	14	10	140	6	4	slider
Erbas et al. (2015) – study 2	95	7	10	70	4	2	slider
Sheets et al. (2015)	30	1	16	16	6	4	Likert
Erbas et al. (2016)	98	7	10	70	4	4	slider
Starr et al. (2017) – study 1	152	5	5	25	7	7	-
Starr et al. (2017) – study 2	50	7	6	42	7	7	Likert
Tong and Keng (2017)	390	2	18	36	8	-	Likert
Lennarz et al. (2018)	86	6	7.33[c]	44	9	5	Likert
Widdershoven et al. (2019)	79	5	10	50	8	4	Likert
Kalokerinos et al. (2019) – study 2	101	7	10	63	6	-	slider
van der Gucht et al. (2019)	61	4	10	40	6	6	slider
Krafft et al. (2020)	60	7	4	28	4	4	Likert
Ottenstein and Lischetzke (2020) – study 2	190	21	1	21	6	-	Likert
Williams-Kerver and Crowther (2020)	118	7	5	35	6	7	Likert

Table A1 (continued)

Source	$N_{\text{Level 2}}$	N_{days}	$N_{\text{occ per day}}$	$N_{\text{occ total}}$	$N_{\text{items negative}}$	$N_{\text{items positive}}$	Response scale
Willroth et al. (2020)	220	14	1	14	11	5	Likert
Starr et al. (2020)	233	7	4	28	12	5	Likert
Oh and Tong (2020)	1,010	8	1	8	14	-	Likert
Liu et al. (2020)	65	7	8	56	7	6	Likert
Seah et al. (2020) – study 1	54	14	5	70	6	-	Likert
Seah et al. (2020) – study 2	193	10	5	50	6	-	Likert
MacCann et al. (2020) – study 2	122	5	-	24	6	3	Likert
Mikhail et al. (2020)	482	45	1	45	10	10	Likert
D'Mello and Gruber (2021)	581	56	1	56	5	5	Likert
Hoemann et al. (2021)	50	14	1	14	8	10	Likert
Brown et al. (2021)	502	30[d]	1[d]	30[d]	8	-	Likert
Lischetzke et al. (2021)	313	21	1	21	8	-	slider
Thompson et al. (2021)	215	14	5	70	6	6	Likert
Lazarus and Fisher (2021)	32	30	4	120	7	4	slider

Table A1 (continued)

Source	$N_{\text{Level 2}}$	N_{days}	$N_{\text{occ per day}}$	$N_{\text{occ total}}$	$N_{\text{items negative}}$	$N_{\text{items positive}}$	Response scale
Bicaker et al. (2022)	192	10	7	70	23	-	Likert
Eckland et al. (2022) – sample 1	277	7	1	7	5	5	Likert
Eckland et al. (2022) – sample 2	140	7	1	7	4	4	Likert
Eckland et al. (2022) – sample 3	197	7	1	7	5	5	Likert
Yang (2022)	248	7	1	7	5	5	Likert
Emery et al. (2022)	102	6	3	18	5	3	Likert
Dong and Xu (2022) – study 1	61	30	1	30	5	5	Likert
Dong and Xu (2022) – study 2	46	3	1	3	10	10	-
Schreuder et al. (2022)	839	6	10	60	8	-	Likert
Pugach et al. (2023)	80	3	17	51	10	-	Likert
Hoemann et al. (2023) – study 2	76	4	10	40	20	19	Likert
Hoemann et al. (2023) – study 3	49	14	10	140	5	2	slider
Seah et al. (2023)	353	10	1	10	6	6	Likert
Walukevich-Dienst et al. (2023)	409	84	1	84	5	-	Likert

Table A1 (continued)

Source	$N_{\text{Level 2}}$	N_{days}	$N_{\text{occ per day}}$	$N_{\text{occ total}}$	$N_{\text{items negative}}$	$N_{\text{items positive}}$	Response scale
Hohensee et al. (2024)	172	28	1	28	7	5	Likert
Liao et al. (2024)	101	7	6	42	8	-	Likert
Seah and Coifman (2024)	83	7	8	56	6	6	Likert
Sels et al. (2024) – study 1	114	7	1	7	8	-	Likert
Sels et al. (2024) – study 4	209	7	10	70	4	4	slider
Knapp et al. (2024)	181	84	1	84	10	10	Likert
Yue et al. (2024)	30	10	5	50	10	-	Likert

Note. Duplicates were removed from the analysis. Studies were only included if emotion differentiation was assessed at the between-person level (i.e., wave-level, day-level or momentary emotion differentiation were excluded). $N_{\text{Level 2}}$ = number of participants in sample. N_{days} = number of days assessed. $N_{\text{occ per day}}$ = number of measurement occasions per day. $N_{\text{occ total}}$ = total number of measurement occasions assessed. $N_{\text{items negative}}$ = number of negative emotion items assessed. $N_{\text{items positive}}$ = number of positive emotion items assessed.

[a] Number refers to the mean number of occasions per participant, as the social-interaction occasions were self-prompted.

[b] The design included self-initiated prompts, such that these values reflect the average number of occasions (per day or total) across participants.

[c] Participants completed two weekends with 22 prompts each. Of these, four prompts were on Friday, and 9 were on each of the following two days. Thus, the average number of prompts per day was 7.33.

[d] There was a total of four measurement waves. Intraclass correlations were calculated for each wave, and then averaged across waves to obtain the final emotion differentiation scores. Thus, we report the number of occasions per wave that were used to calculate the intraclass correlations. We did not exclude this study because emotion differentiation was ultimately examined at the between-person level (i.e., scores were averaged across waves to obtain a between-person measure of emotion differentiation).

B: Descriptive Statistics of Emotion Variables in the Primary Data Set

Table B1

Descriptive Statistics of Emotion Variables in the Primary Data Set

Variable	*M*	SD_{within}	$SD_{between}$	ICC	Range
Negative emotions					
angry	16.63	17.98	13.47	.36	0 – 100
irritated	19.57	17.87	16.69	.47	0 – 100
annoyed	23.88	20.41	17.20	.42	0 – 100
sad	19.85	18.35	17.07	.46	0 – 100
downhearted	22.58	19.96	17.16	.42	0 – 100
unhappy	25.35	18.98	18.16	.48	0 – 100
terrified	9.22	12.40	11.56	.46	0 – 100
fearful	12.97	14.66	15.16	.52	0 – 100
worried	25.10	20.24	17.79	.44	0 – 100
ashamed	9.48	12.63	11.00	.43	0 – 100
humiliated	8.38	11.83	10.94	.46	0 – 100
disgraced	6.92	9.05	9.65	.53	0 – 100
guilty	11.38	13.79	11.61	.41	0 – 100
repentant	14.53	17.16	12.99	.36	0 – 100
I had a bad conscience	14.40	16.29	13.20	.40	0 – 100
Positive emotions					
joyful	43.50	21.17	17.11	.40	0 – 100
glad	47.55	18.67	18.70	.50	0 – 100
happy	52.79	17.45	17.98	.52	0 – 100
interested	47.32	19.37	17.90	.46	0 – 100
alert	55.76	19.51	14.67	.36	0 – 100
curious	41.84	19.95	18.17	.45	0 – 100
love	40.75	20.76	20.79	.50	0 – 100
closeness	41.88	22.18	20.11	.45	0 – 100
trust	40.58	18.47	21.59	.58	0 – 100
proud	32.80	18.20	21.74	.59	0 – 100
confident	46.60	16.89	21.09	.61	0 – 100
self-assured	45.86	16.54	21.18	.62	0 – 100

Note. *M* = grand mean. SD_{within} = within-person standard deviation. $SD_{between}$ = between-person standard deviation. ICC = intraclass correlation.

C: Results Tables and Additional Figures for the Main Analyses in the Primary Data Set

Table C1

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Negative Emotion Differentiation in the Primary Data Set)

Condition		MCSE					
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC
14	5	0.001	0.004	0.002	0.001	0.000	0.000
14	10	0.001	0.004	0.002	0.001	0.000	0.000
14	15	0.001	0.004	0.003	0.001	0.000	0.000
20	5	0.001	0.004	0.002	0.001	0.000	0.000
20	10	0.001	0.004	0.002	0.001	0.000	0.000
20	15	0.001	0.004	0.002	0.000	0.000	0.000
30	5	0.001	0.004	0.002	0.001	0.000	0.000
30	10	0.001	0.004	0.002	0.000	0.000	0.000
30	15	0.001	0.004	0.002	0.000	0.000	0.000
40	5	0.001	0.004	0.003	0.001	0.000	0.000
40	10	0.001	0.004	0.003	0.000	0.000	0.000
40	15	0.001	0.004	0.003	0.000	0.000	0.000
50	5	0.001	0.004	0.003	0.000	0.000	0.000
50	10	0.001	0.004	0.003	0.000	0.000	0.000
50	15	0.001	0.003	0.003	0.000	0.000	0.000
60	5	0.001	0.003	0.003	0.000	0.000	0.000
60	10	0.000	0.003	0.003	0.000	0.000	0.000
60	15	0.000	0.003	0.003	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero.

Table C2

Results from the Simulation Study for Raw Intraclass Correlations (ICCs) Aggregated Across Replications (Negative Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	5	.254	.562	.780	-0.409	-0.056	0.169	0.084	0.148	0.496	0.128	0.173	0.216	.097	.461	.713	.009	.098	.229
14	10	.402	.653	.867	-0.400	-0.027	0.096	0.046	0.108	0.483	0.118	0.151	0.184	.339	.585	.732	.000	.022	.083
14	15	.463	.685	.851	-0.398	-0.019	0.056	0.034	0.096	0.478	0.116	0.145	0.175	.444	.632	.758	.000	.009	.046
20	5	.384	.633	.825	-0.364	-0.051	0.139	0.075	0.129	0.472	0.123	0.160	0.201	.280	.572	.761	.000	.068	.165
20	10	.514	.726	.883	-0.350	-0.021	0.072	0.037	0.090	0.454	0.113	0.141	0.173	.472	.674	.795	.000	.010	.064
20	15	.539	.760	.901	-0.335	-0.013	0.034	0.027	0.077	0.443	0.112	0.136	0.160	.570	.709	.800	.000	.003	.028
30	5	.473	.716	.884	-0.283	-0.046	0.109	0.065	0.109	0.423	0.117	0.149	0.187	.486	.674	.811	.000	.043	.128
30	10	.645	.809	.924	-0.266	-0.016	0.046	0.030	0.070	0.403	0.107	0.132	0.157	.634	.753	.826	.000	.003	.037
30	15	.679	.842	.941	-0.262	-0.008	0.015	0.019	0.058	0.395	0.107	0.127	0.146	.683	.779	.840	.000	.000	.018
40	5	.553	.774	.911	-0.204	-0.043	0.090	0.058	0.097	0.368	0.115	0.143	0.174	.607	.735	.833	.000	.030	.110
40	10	.708	.868	.950	-0.189	-0.014	0.027	0.024	0.057	0.346	0.105	0.127	0.145	.708	.799	.846	.000	.001	.028
40	15	.772	.900	.963	-0.186	-0.005	0.009	0.014	0.044	0.339	0.105	0.122	0.136	.761	.820	.861	.000	.000	.009
50	5	.622	.819	.926	-0.135	-0.041	0.070	0.055	0.088	0.307	0.113	0.138	0.170	.674	.774	.858	.000	.022	.092
50	10	.761	.913	.968	-0.118	-0.011	0.019	0.021	0.047	0.279	0.105	0.123	0.139	.764	.829	.869	.000	.000	.018
50	15	.810	.945	.982	-0.116	-0.003	0.004	0.010	0.032	0.272	0.105	0.119	0.131	.807	.847	.877	.000	.000	.000
60	5	.637	.853	.928	-0.104	-0.040	0.059	0.048	0.080	0.239	0.114	0.135	0.163	.737	.803	.876	.000	.018	.083
60	10	.807	.946	.977	-0.062	-0.010	0.013	0.018	0.038	0.203	0.106	0.121	0.133	.813	.851	.879	.000	.000	.009
60	15	.853	.978	.994	-0.054	-0.001	0.002	0.007	0.021	0.190	0.107	0.117	0.125	.844	.866	.884	.000	.000	.000
70	15	-	-	-	-	-	-	-	-	-	-	0.115	-	-	.881	-	-	.000	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 15 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The number of participants that were used for the reliability analysis was 109 across all conditions and replications. There were no upper bound (ICCs $\geq$ 1) or lower bound (ICCs $\leq$ -1/[k-1]) estimation problems in any condition. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criteria, they reflect the minimum and maximum value across replications. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table C3

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Positive Emotion Differentiation in the Primary Data Set)

Condition		MCSE					
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC
14	4	0.001	0.003	0.003	0.001	0.000	0.000
14	8	0.001	0.002	0.002	0.001	0.000	0.000
14	12	0.001	0.002	0.002	0.001	0.000	0.000
20	4	0.001	0.003	0.002	0.001	0.000	0.000
20	8	0.000	0.002	0.002	0.001	0.000	0.000
20	12	0.000	0.002	0.002	0.001	0.000	0.000
30	4	0.001	0.003	0.002	0.001	0.000	0.000
30	8	0.000	0.002	0.001	0.000	0.000	0.000
30	12	0.000	0.001	0.001	0.000	0.000	0.000
40	4	0.000	0.003	0.002	0.001	0.000	0.000
40	8	0.000	0.002	0.001	0.000	0.000	0.000
40	12	0.000	0.001	0.001	0.000	0.000	0.000
50	4	0.000	0.003	0.002	0.000	0.000	0.000
50	8	0.000	0.001	0.001	0.000	0.000	0.000
50	12	0.000	0.001	0.001	0.000	0.000	0.000
60	4	0.000	0.003	0.002	0.000	0.000	0.000
60	8	0.000	0.001	0.001	0.000	0.000	0.000
60	12	0.000	0.000	0.000	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero.

Table C4

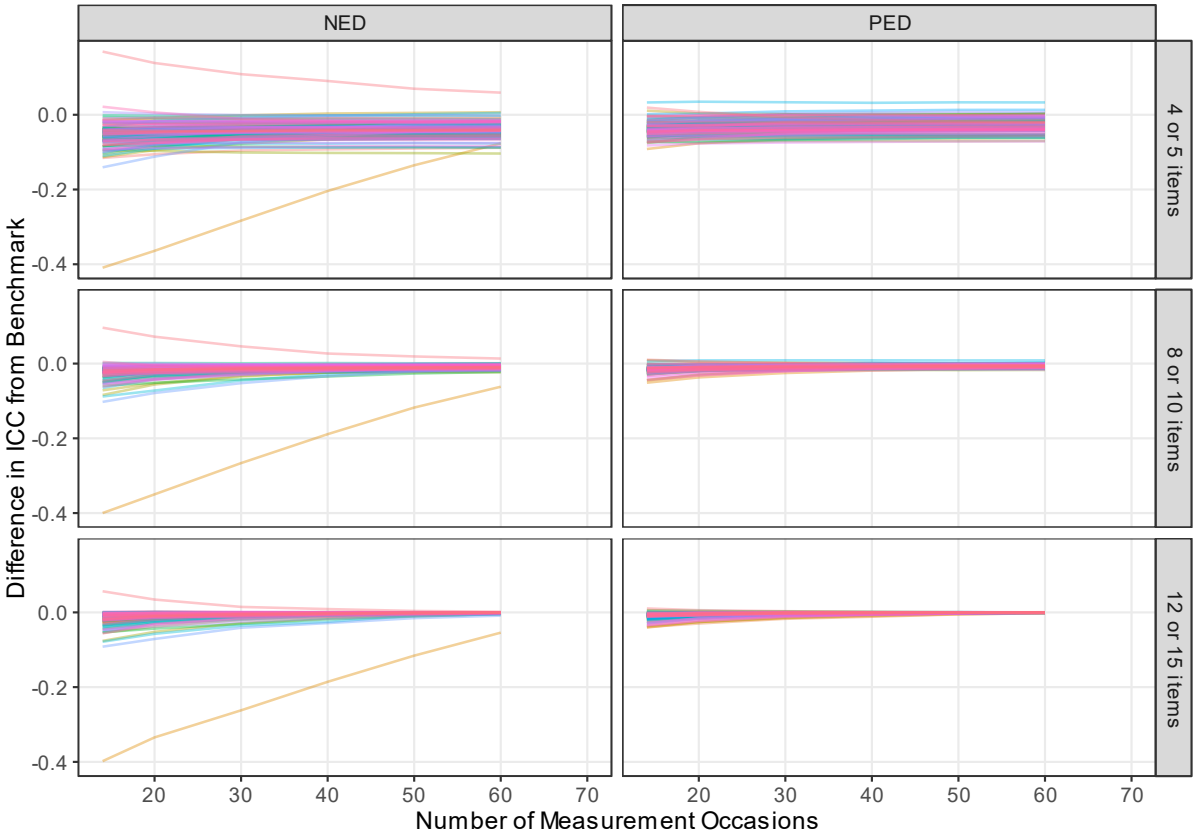
Results from the Simulation Study for Raw Intraclass Correlations (ICCs) Aggregated Across Replications (Positive Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	4	.407	.622	.805	-0.092	-0.038	0.033	0.106	0.160	0.235	0.152	0.195	0.239	.052	.449	.682	.000	.055	.165
14	8	.505	.731	.843	-0.051	-0.016	0.010	0.074	0.113	0.173	0.131	0.163	0.197	.298	.532	.700	.000	.010	.055
14	12	.591	.769	.881	-0.041	-0.010	0.011	0.056	0.097	0.164	0.128	0.154	0.181	.340	.561	.702	.000	.003	.028
20	4	.423	.692	.816	-0.077	-0.034	0.035	0.086	0.136	0.217	0.135	0.178	0.218	.173	.545	.721	.000	.035	.119
20	8	.661	.801	.895	-0.037	-0.012	0.009	0.059	0.092	0.148	0.121	0.150	0.180	.397	.622	.748	.000	.004	.037
20	12	.715	.839	.916	-0.030	-0.006	0.006	0.043	0.076	0.125	0.118	0.142	0.164	.439	.647	.755	.000	.001	.028
30	4	.617	.762	.867	-0.074	-0.030	0.033	0.068	0.113	0.197	0.128	0.163	0.197	.402	.644	.765	.000	.020	.083
30	8	.777	.869	.930	-0.025	-0.009	0.009	0.047	0.071	0.128	0.113	0.140	0.163	.558	.711	.789	.000	.002	.028
30	12	.819	.906	.946	-0.017	-0.003	0.003	0.031	0.055	0.094	0.115	0.133	0.149	.633	.733	.796	.000	.000	.009
40	4	.658	.806	.897	-0.073	-0.029	0.032	0.056	0.100	0.190	0.123	0.155	0.185	.523	.705	.808	.000	.014	.055
40	8	.837	.909	.951	-0.019	-0.008	0.009	0.036	0.058	0.115	0.117	0.134	0.155	.685	.766	.828	.000	.001	.018
40	12	.896	.944	.969	-0.012	-0.002	0.001	0.023	0.042	0.071	0.115	0.128	0.142	.727	.786	.827	.000	.000	.009
50	4	.704	.836	.911	-0.071	-0.028	0.033	0.050	0.091	0.187	0.122	0.150	0.177	.623	.749	.829	.000	.010	.055
50	8	.882	.936	.968	-0.018	-0.007	0.008	0.029	0.049	0.103	0.115	0.131	0.153	.748	.803	.857	.000	.000	.009
50	12	.941	.969	.983	-0.005	-0.001	0.001	0.017	0.030	0.052	0.115	0.125	0.135	.787	.820	.852	.000	.000	.009
60	4	.747	.858	.924	-0.071	-0.028	0.033	0.045	0.084	0.185	0.118	0.146	0.168	.667	.781	.841	.000	.008	.046
60	8	.913	.955	.978	-0.017	-0.006	0.009	0.023	0.041	0.096	0.113	0.128	0.144	.783	.830	.864	.000	.000	.009
60	12	.976	.987	.993	-0.002	0.000	0.000	0.011	0.019	0.034	0.117	0.123	0.130	.828	.845	.863	.000	.000	.000
70	12	-	-	-	-	-	-	-	-	-	-	0.121	-	-	.864	-	-	.000	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 12 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The number of participants that were used for the reliability analysis was 109 across all conditions and replications. There were no upper bound (ICCs ≥ 1) or lower bound (ICCs ≤ -1/[k-1]) estimation problems in any condition. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criteria, they reflect the minimum and maximum value across replications. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Figure C1

Person-Level Difference from Raw Benchmark Intraclass Correlations (ICC) Separate for Each Participant



Note. Each line represents a participant's person-level difference between the ICC in a given simulation condition and the ICC in the benchmark condition. The left column shows person-level differences negative emotion differentiation (NED), whereas the right column shows the person-level differences for positive emotion differentiation (PED). The top, middle, and bottom rows display results for conditions using 5, 10, and 15 emotion items for NED, or 4, 8, and 12 items for PED, respectively.

D: Sensitivity Analysis Using Fisher's *Z*-transformed Intraclass Correlations in the Primary Data Set

Table D1

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Fisher's Z-Transformed Intraclass Correlations (ICCs) Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Negative Emotion Differentiation in the Primary Data Set)

Condition		MCSE					
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC
14	5	0.001	0.008	0.006	0.001	0.000	0.000
14	10	0.001	0.008	0.007	0.001	0.000	0.000
14	15	0.001	0.009	0.007	0.001	0.000	0.000
20	5	0.001	0.008	0.006	0.001	0.000	0.000
20	10	0.001	0.009	0.005	0.001	0.000	0.000
20	15	0.001	0.009	0.006	0.000	0.000	0.000
30	5	0.001	0.008	0.006	0.001	0.000	0.000
30	10	0.001	0.009	0.006	0.000	0.000	0.000
30	15	0.001	0.009	0.005	0.000	0.000	0.000
40	5	0.001	0.008	0.007	0.001	0.000	0.000
40	10	0.001	0.009	0.007	0.000	0.000	0.000
40	15	0.001	0.009	0.006	0.000	0.000	0.000
50	5	0.001	0.007	0.007	0.000	0.000	0.000
50	10	0.001	0.007	0.007	0.000	0.000	0.000
50	15	0.001	0.008	0.006	0.000	0.000	0.000
60	5	0.001	0.006	0.006	0.000	0.000	0.000
60	10	0.000	0.006	0.007	0.000	0.000	0.000
60	15	0.000	0.006	0.007	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero.

Table D2

Results from the Simulation Study for Fisher's Z-Transformed Intraclass Correlations (ICCs) Aggregated Across Replications (Negative Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	5	.252	.557	.774	-1.211	-0.504	-0.097	0.354	0.564	1.328	0.241	0.315	0.428	.097	.461	.713	.009	.098	.229
14	10	.407	.651	.850	-1.033	-0.236	0.003	0.218	0.334	1.195	0.266	0.338	0.417	.339	.585	.732	.000	.022	.083
14	15	.378	.680	.848	-0.908	-0.065	0.079	0.112	0.241	1.104	0.285	0.352	0.431	.444	.632	.758	.000	.009	.046
20	5	.382	.630	.818	-1.124	-0.493	-0.151	0.337	0.538	1.263	0.221	0.289	0.383	.280	.572	.761	.000	.068	.165
20	10	.487	.724	.879	-0.918	-0.217	-0.037	0.190	0.292	1.110	0.243	0.311	0.390	.472	.674	.795	.000	.010	.064
20	15	.552	.756	.897	-0.748	-0.044	0.043	0.084	0.190	0.998	0.264	0.322	0.387	.570	.709	.800	.000	.003	.028
30	5	.481	.711	.874	-0.977	-0.482	-0.201	0.317	0.514	1.141	0.209	0.265	0.346	.486	.674	.811	.000	.043	.128
30	10	.640	.805	.918	-0.738	-0.202	-0.079	0.173	0.254	0.976	0.234	0.286	0.339	.634	.753	.826	.000	.003	.037
30	15	.690	.838	.937	-0.575	-0.026	0.014	0.060	0.140	0.871	0.247	0.296	0.348	.683	.779	.840	.000	.000	.018
40	5	.540	.768	.903	-0.836	-0.475	-0.230	0.311	0.501	1.013	0.206	0.252	0.315	.607	.735	.833	.000	.030	.110
40	10	.704	.864	.944	-0.576	-0.193	-0.111	0.160	0.230	0.833	0.225	0.271	0.309	.708	.799	.846	.000	.001	.028
40	15	.742	.895	.961	-0.405	-0.016	0.010	0.044	0.106	0.739	0.244	0.281	0.319	.761	.820	.861	.000	.000	.009
50	5	.632	.811	.922	-0.718	-0.471	-0.257	0.306	0.491	0.882	0.201	0.243	0.305	.674	.774	.858	.000	.022	.092
50	10	.755	.909	.967	-0.429	-0.186	-0.114	0.152	0.214	0.675	0.222	0.262	0.298	.764	.829	.869	.000	.000	.018
50	15	.804	.942	.982	-0.251	-0.009	0.003	0.031	0.077	0.587	0.242	0.271	0.302	.807	.847	.877	.000	.000	.000
60	5	.646	.845	.917	-0.618	-0.468	-0.256	0.302	0.485	0.748	0.203	0.237	0.296	.737	.803	.876	.000	.018	.083
60	10	.802	.943	.974	-0.315	-0.182	-0.112	0.146	0.201	0.507	0.227	0.255	0.282	.813	.851	.879	.000	.000	.009
60	15	.843	.976	.993	-0.118	-0.004	0.002	0.019	0.050	0.408	0.245	0.264	0.284	.844	.866	.884	.000	.000	.000
70	15	-	-	-	-	-	-	-	-	-	-	0.259	-	-	.881	-	-	.000	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 15 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The number of participants that were used for the reliability analysis was 109 across all conditions and replications. There were no upper bound (ICCs ≥ 1) or lower bound (ICCs ≤ -1/[k-1]) estimation problems in any condition. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criteria, they reflect the minimum and maximum value across replications. The values for the reliability are identical to those reported for the raw ICCs since the reliability estimates are always based on Fisher's *Z*-transformed ICCs, and the values for % Neg ICC are identical because they refer to raw ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table D3

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Fisher's Z-Transformed Intraclass Correlations (ICCs) Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Positive Emotion Differentiation in in the Primary Data Set)

Condition		MCSE					
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC
14	4	0.001	0.006	0.005	0.001	0.000	0.000
14	8	0.001	0.005	0.005	0.001	0.000	0.000
14	12	0.001	0.005	0.005	0.001	0.000	0.000
20	4	0.001	0.005	0.004	0.001	0.000	0.000
20	8	0.000	0.004	0.004	0.001	0.000	0.000
20	12	0.000	0.004	0.003	0.001	0.000	0.000
30	4	0.001	0.005	0.004	0.001	0.000	0.000
30	8	0.000	0.003	0.003	0.000	0.000	0.000
30	12	0.000	0.003	0.002	0.000	0.000	0.000
40	4	0.000	0.004	0.003	0.001	0.000	0.000
40	8	0.000	0.003	0.003	0.000	0.000	0.000
40	12	0.000	0.002	0.002	0.000	0.000	0.000
50	4	0.000	0.004	0.003	0.000	0.000	0.000
50	8	0.000	0.003	0.002	0.000	0.000	0.000
50	12	0.000	0.002	0.002	0.000	0.000	0.000
60	4	0.000	0.004	0.003	0.000	0.000	0.000
60	8	0.000	0.003	0.002	0.000	0.000	0.000
60	12	0.000	0.001	0.001	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero.

Table D4

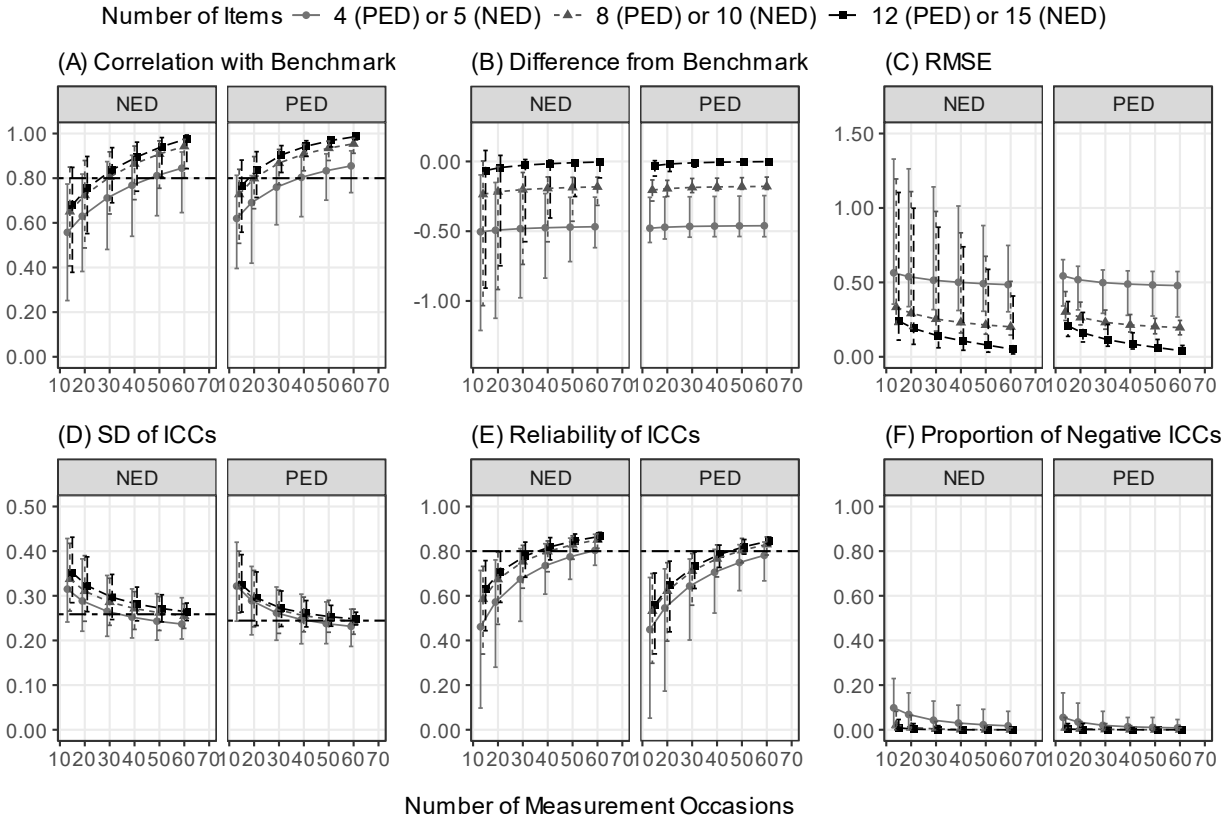
Results from the Simulation Study for Fisher's Z-Transformed Intraclass Correlations (ICCs) Aggregated Across Replications (Positive Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	4	.396	.620	.813	-0.581	-0.479	-0.257	0.341	0.543	0.653	0.243	0.322	0.420	.052	.449	.682	.000	.055	.165
14	8	.508	.728	.840	-0.284	-0.205	-0.133	0.244	0.302	0.438	0.262	0.322	0.400	.298	.532	.700	.000	.010	.055
14	12	.558	.766	.881	-0.105	-0.030	0.006	0.137	0.208	0.371	0.264	0.325	0.392	.340	.561	.702	.000	.003	.028
20	4	.419	.691	.811	-0.556	-0.471	-0.254	0.315	0.518	0.608	0.213	0.289	0.366	.173	.545	.721	.000	.035	.119
20	8	.663	.799	.897	-0.250	-0.195	-0.136	0.214	0.264	0.367	0.230	0.293	0.357	.397	.622	.748	.000	.004	.037
20	12	.714	.837	.919	-0.072	-0.018	0.002	0.100	0.160	0.298	0.233	0.295	0.353	.439	.647	.755	.000	.001	.028
30	4	.591	.760	.861	-0.543	-0.466	-0.253	0.292	0.498	0.584	0.200	0.262	0.320	.402	.644	.765	.000	.020	.083
30	8	.772	.868	.929	-0.223	-0.186	-0.122	0.185	0.231	0.314	0.216	0.268	0.313	.558	.711	.789	.000	.002	.028
30	12	.825	.904	.946	-0.041	-0.009	0.002	0.071	0.114	0.218	0.231	0.272	0.311	.633	.733	.796	.000	.000	.009
40	4	.628	.804	.895	-0.539	-0.463	-0.251	0.279	0.489	0.577	0.193	0.247	0.304	.523	.705	.808	.000	.014	.055
40	8	.833	.908	.951	-0.209	-0.182	-0.120	0.167	0.214	0.284	0.219	0.255	0.297	.685	.766	.828	.000	.001	.018
40	12	.898	.943	.968	-0.027	-0.005	0.001	0.052	0.085	0.161	0.230	0.260	0.289	.727	.786	.827	.000	.000	.009
50	4	.702	.833	.909	-0.538	-0.462	-0.248	0.274	0.483	0.573	0.193	0.238	0.286	.623	.749	.829	.000	.010	.055
50	8	.879	.935	.966	-0.20	-0.180	-0.116	0.155	0.204	0.259	0.218	0.248	0.290	.748	.803	.857	.000	.000	.009
50	12	.940	.969	.983	-0.013	-0.003	0.001	0.038	0.062	0.117	0.232	0.253	0.279	.787	.820	.852	.000	.000	.009
60	4	.735	.855	.923	-0.54	-0.461	-0.245	0.268	0.479	0.573	0.187	0.232	0.270	.667	.781	.841	.000	.008	.046
60	8	.912	.954	.979	-0.201	-0.179	-0.112	0.148	0.196	0.244	0.214	0.243	0.270	.783	.830	.864	.000	.000	.009
60	12	.972	.987	.993	-0.006	-0.001	0.000	0.025	0.040	0.076	0.235	0.248	0.263	.828	.845	.863	.000	.000	.000
70	12	-	-	-	-	-	-	-	-	-	-	0.244	-	-	.864	-	-	.000	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 15 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The number of participants that were used for the reliability analysis was 109 across all conditions and replications. There were no upper bound (ICCs ≥ 1) or lower bound (ICCs ≤ -1/[k-1]) estimation problems in any condition. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criteria, they reflect the minimum and maximum value across replications. The values for the reliability are identical to those reported for the raw ICCs since the reliability estimates are always based on Fisher's *Z*-transformed ICCs, and the values for % Neg ICC are identical because they refer to raw ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Figure D1

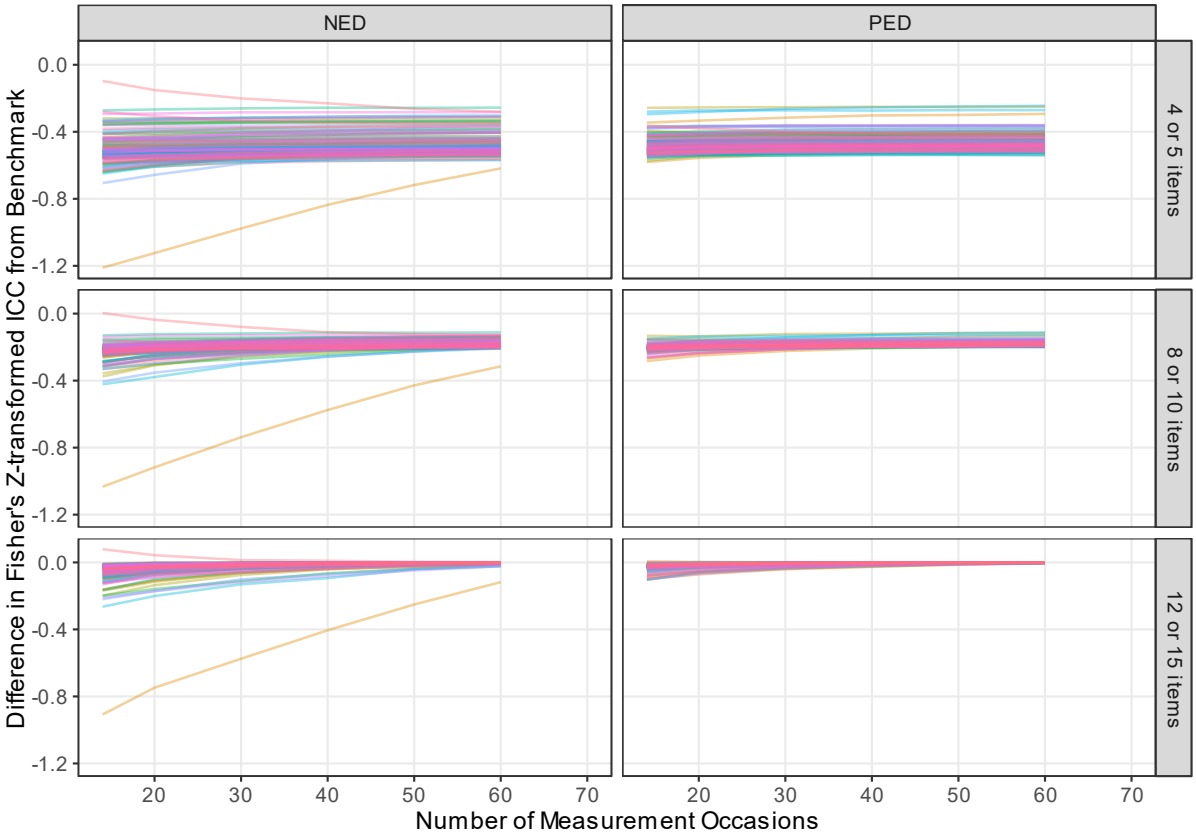
Results of the Evaluation Criteria for Fisher's Z-Transformed Intraclass Correlations (ICCs) for Each Condition



Note. Panels (A) to (F) depict six evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (F) are absolute criteria. The panels present results for negative emotion differentiation (NED; left) and positive emotion differentiation (PED; right). The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across 5,000 replications. The plots display mean values for each item number condition. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E), and the standard deviation (*SD*) of ICCs in the benchmark condition in (D). Proportion of negative ICCs uses raw ICCs; all other criteria use Fisher's *Z*-transformed ICCs.

Figure D2

Person-Level Difference from Fisher's Z-Transformed Benchmark Intraclass Correlations (ICC) Separate for Each Participant



Note. Each line represents a participant's person-level difference between the Fisher's *Z*-transformed ICC in a given simulation condition and the Fisher's *Z*-transformed ICC in the benchmark condition. The left column shows person-level differences for negative emotion differentiation (NED), whereas the right column shows the person-level differences for positive emotion differentiation (PED). The top, middle, and bottom rows display results for conditions using 5, 10, and 15 emotion items for NED, or 4, 8, and 12 items for PED, respectively.

E: Sensitivity Analyses Using Ordered Draws of Measurement Occasions in the Primary Data Set

Table E1

Results from the Sensitivity Analysis for Raw Intraclass Correlations (ICCs) Using Measurement Occasions Drawn by Temporal Order (Negative Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			*SD*			Reliability			% Neg ICC		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	5	.334	.476	.595	-0.499	-0.048	0.317	0.162	0.181	0.203	.394	.509	.605	.028	.090	.156
14	10	.477	.540	.596	-0.494	-0.019	0.318	0.150	0.158	0.165	.555	.607	.649	.000	.015	.046
14	15	-	.558	-	-0.493	-0.011	0.319	-	0.152	-	-	.647	-	-	.009	-
20	5	.437	.578	.692	-0.480	-0.040	0.339	0.150	0.168	0.190	.519	.611	.694	.009	.059	.138
20	10	.597	.654	.698	-0.476	-0.009	0.342	0.140	0.147	0.154	.636	.679	.714	.000	.004	.028
20	15	-	.678	-	-0.475	0.000	0.342	-	0.140	-	-	.701	-	-	.000	-
30	5	.522	.647	.729	-0.502	-0.037	0.214	0.140	0.156	0.181	.621	.699	.778	.000	.041	.101
30	10	.682	.729	.755	-0.501	-0.007	0.232	0.126	0.136	0.144	.707	.761	.792	.000	.001	.018
30	15	-	.756	-	-0.501	0.001	0.238	-	0.131	-	-	.783	-	-	.000	-
40	5	.575	.695	.780	-0.531	-0.040	0.180	0.129	0.143	0.163	.673	.737	.795	.000	.031	.083
40	10	.743	.774	.810	-0.532	-0.010	0.200	0.118	0.127	0.133	.760	.799	.817	.000	.000	.009
40	15	-	.800	-	-0.532	-0.001	0.205	-	0.123	-	-	.820	-	-	.000	-
50	5	.579	.728	.810	-0.541	-0.044	0.103	0.121	0.136	0.156	.709	.769	.824	.000	.027	.064
50	10	.772	.815	.847	-0.509	-0.012	0.119	0.112	0.120	0.126	.793	.824	.841	.000	.000	.009
50	15	-	.846	-	-0.492	-0.003	0.127	-	0.116	-	-	.842	-	-	.000	-
60	5	.635	.775	.844	-0.546	-0.044	0.042	0.116	0.130	0.150	.741	.790	.840	.000	.022	.055
60	10	.823	.860	.888	-0.519	-0.014	0.064	0.108	0.116	0.123	.817	.844	.862	.000	.000	.000
60	15	-	.890	-	-0.503	-0.005	0.072	-	0.113	-	-	.861	-	-	.000	-
70	5	.780	.882	.932	-0.103	-0.039	0.050	0.119	0.132	0.155	.788	.826	.880	.000	.014	.055
70	10	.947	.972	.984	-0.023	-0.009	0.010	0.109	0.119	0.125	.845	.867	.880	.000	.000	.000
70	15	-	-	-	-	-	-	-	0.115	-	-	.881	-	-	.000	-

Note. The last row reflects the benchmark condition. The number of participants that were used for the reliability analysis was 109 across all conditions and replications. There were no upper bound (ICCs ≥ 1) or lower bound (ICCs ≤ -1/[k-1]) estimation problems in any condition. For the ordered draws of measurement occasions, we examined all possible item sets for each condition (with one replication each); thus, the minimum and maximum values reflect minimum and maximum values across the possible item sets. For the difference in ICCs, these values reflect the minimum and maximum value of the person-level average difference in ICCs across item sets. Since there was only one item set containing the maximum number of items, minimum and maximum values do not apply here. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table E2

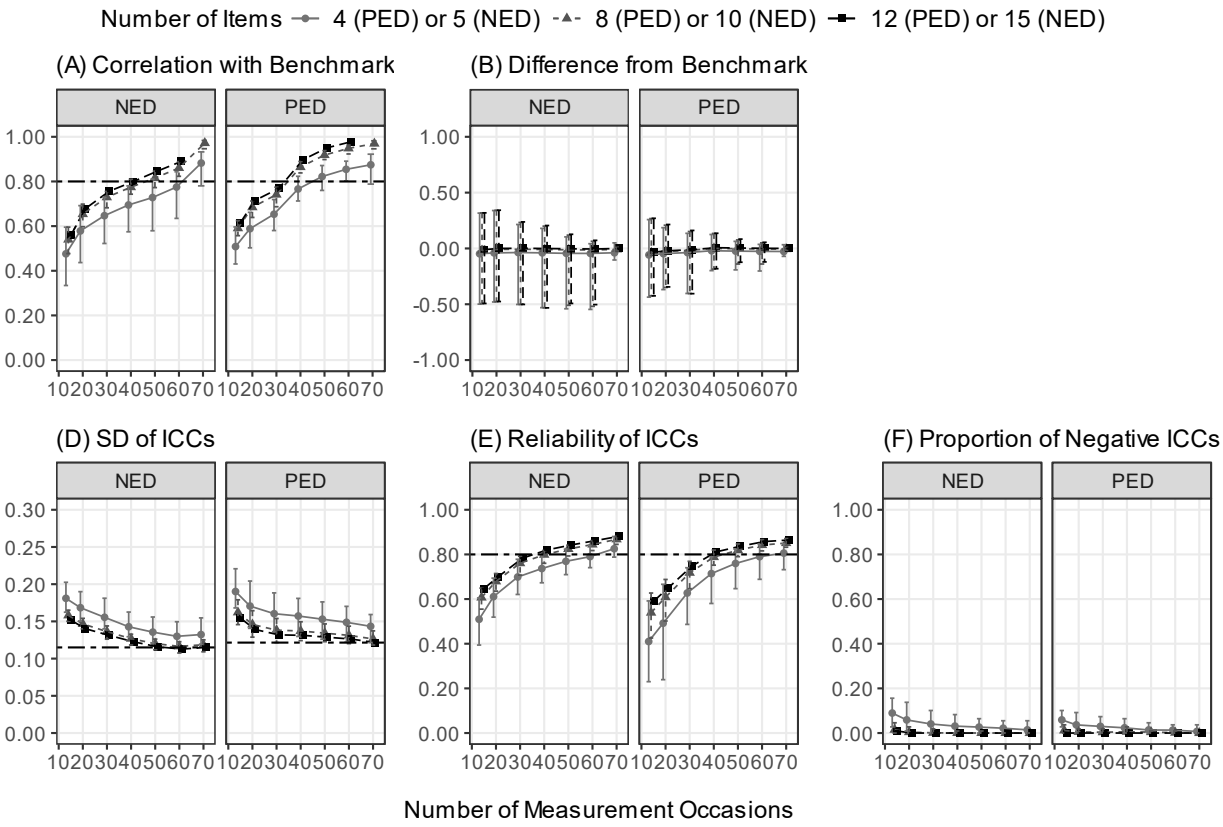
Results from the Sensitivity Analysis for Raw Intraclass Correlations (ICCs) Using Measurement Occasions Drawn by Temporal Order (Positive Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			*SD*			Reliability			% Neg ICC		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	4	.430	.508	.596	-0.436	-0.058	0.258	0.168	0.190	0.221	.230	.410	.592	.028	.059	.101
14	8	.557	.590	.625	-0.425	-0.038	0.267	0.146	0.162	0.179	.426	.539	.627	.000	.012	.037
14	12	-	.615	-	-0.425	-0.033	0.270	-	0.155	-	-	.590	-	-	.000	-
20	4	.503	.588	.662	-0.369	-0.047	0.186	0.141	0.171	0.204	.239	.492	.666	.000	.037	.092
20	8	.639	.684	.712	-0.349	-0.027	0.208	0.129	0.146	0.164	.502	.608	.688	.000	.005	.018
20	12	-	.714	-	-0.345	-0.021	0.214	-	0.139	-	-	.649	-	-	.000	-
30	4	.580	.653	.719	-0.403	-0.037	0.135	0.136	0.160	0.188	.487	.627	.740	.000	.030	.073
30	8	.687	.741	.774	-0.404	-0.017	0.152	0.122	0.138	0.154	.647	.718	.769	.000	.004	.018
30	12	-	.769	-	-0.405	-0.011	0.161	-	0.132	-	-	.748	-	-	.000	-
40	4	.713	.766	.823	-0.199	-0.021	0.125	0.130	0.157	0.181	.580	.714	.789	.000	.023	.064
40	8	.838	.864	.890	-0.185	0.000	0.134	0.124	0.137	0.149	.752	.788	.820	.000	.006	.028
40	12	-	.897	-	-0.182	0.006	0.136	-	0.131	-	-	.811	-	-	.000	-
50	4	.760	.822	.871	-0.192	-0.024	0.063	0.126	0.153	0.176	.647	.759	.824	.000	.014	.046
50	8	.898	.918	.938	-0.140	-0.002	0.076	0.123	0.134	0.147	.790	.819	.847	.000	.004	.009
50	12	-	.951	-	-0.124	0.004	0.084	-	0.129	-	-	.837	-	-	.000	-
60	4	.802	.854	.891	-0.202	-0.026	0.036	0.122	0.149	0.170	.688	.790	.845	.000	.013	.037
60	8	.923	.948	.963	-0.140	-0.005	0.050	0.120	0.131	0.144	.816	.842	.866	.000	.001	.009
60	12	-	.978	-	-0.121	0.001	0.055	-	0.126	-	-	.858	-	-	.000	-
70	4	.788	.875	.922	-0.071	-0.027	0.033	0.121	0.143	0.159	.732	.806	.848	.000	.007	.037
70	8	.946	.970	.982	-0.017	-0.006	0.009	0.118	0.126	0.137	.832	.851	.871	.000	.000	.000
70	12	-	-	-	-	-	-	-	0.121	-	-	.864	-	-	.000	-

Note. The last row reflects the benchmark condition. The number of participants that were used for the reliability analysis was 109 across all conditions and replications. There were no upper bound (ICCs ≥ 1) or lower bound (ICCs ≤ -1/[k-1]) estimation problems in any condition. For the ordered draws of measurement occasions, we examined all possible item sets for each condition (with one replication each); thus, the minimum and maximum values reflect minimum and maximum values across the possible item sets. For the difference in ICCs, these values reflect the minimum and maximum value of the person-level average difference in ICCs across item sets. Since there was only one item set containing the maximum number of items, minimum and maximum values do not apply here. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Figure E1

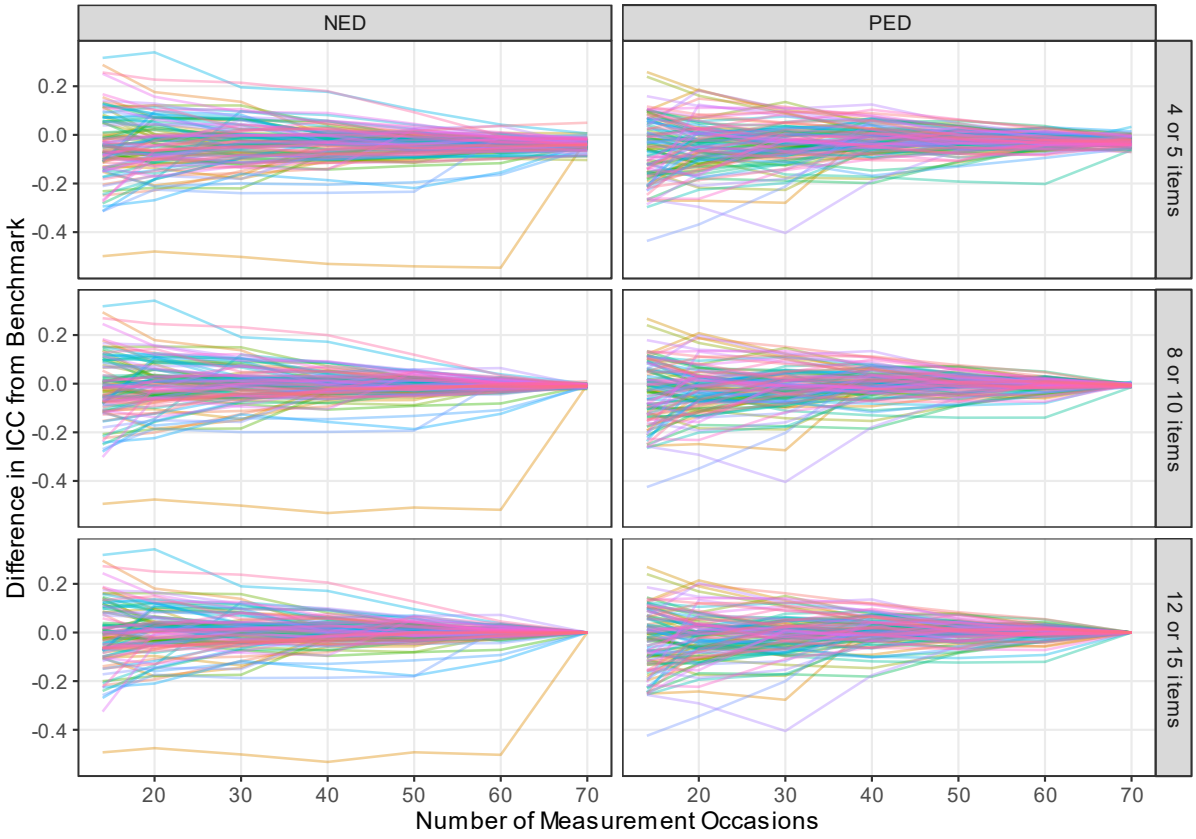
Results of the Evaluation Criteria for raw Intraclass Correlations (ICCs) for Each Condition Using Measurement Occasions Drawn by Temporal Order



Note. Panels (A) to (F) depict the evaluation criteria: (A) to (B) are relative to the benchmark condition; (D) to (F) are absolute criteria. The panels present results for negative emotion differentiation (NED; left) and positive emotion differentiation (PED; right). For the ordered draws of measurement occasions, we examined all possible item sets for each condition (with one replication each); thus, the minimum and maximum values reflect minimum and maximum values across the possible item sets. For the difference in ICCs, these values reflect the minimum and maximum value of the person-level average difference in ICCs across item sets. Since there was only one item set containing the maximum number of items, minimum and maximum values do not apply here. The RMSE was calculated per participant across replications and, thus, does not apply to the ordered draws of measurement occasions. The reliability calculation required Fisher's *Z*-transformed ICCs.

Figure E2

Person-Level Difference from Raw Benchmark Intraclass Correlations (ICC) Separate for Each Participant Using Measurement Occasions Drawn by Temporal Order



Note. Each line represents a participant's person-level difference between the ICC in a given simulation condition and the ICC in the benchmark condition. The left column shows person-level differences negative emotion differentiation (NED), whereas the right column shows the person-level differences for positive emotion differentiation (PED). The top, middle, and bottom rows display results for conditions using 5, 10, and 15 emotion items for NED, or 4, 8, and 12 items for PED, respectively.

F: Sensitivity Analyses Using Different Treatments of Negative Intraclass Correlations in the Primary Data Set

Table F1

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Setting Negative ICCs to Zero Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Negative Emotion Differentiation in the Primary Data Set)

Condition		MCSE					
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC
14	5	0.001	0.004	0.002	0.001	0.000	0.000
14	10	0.001	0.004	0.002	0.001	0.000	0.000
14	15	0.001	0.004	0.003	0.001	0.000	0.000
20	5	0.001	0.004	0.002	0.001	0.000	0.000
20	10	0.001	0.004	0.002	0.001	0.000	0.000
20	15	0.001	0.004	0.002	0.000	0.000	0.000
30	5	0.001	0.004	0.002	0.001	0.000	0.000
30	10	0.001	0.004	0.002	0.000	0.000	0.000
30	15	0.001	0.004	0.002	0.000	0.000	0.000
40	5	0.001	0.004	0.003	0.001	0.000	0.000
40	10	0.001	0.004	0.003	0.000	0.000	0.000
40	15	0.001	0.004	0.003	0.000	0.000	0.000
50	5	0.001	0.004	0.003	0.000	0.000	0.000
50	10	0.001	0.004	0.003	0.000	0.000	0.000
50	15	0.001	0.004	0.003	0.000	0.000	0.000
60	5	0.001	0.003	0.003	0.000	0.000	0.000
60	10	0.000	0.003	0.003	0.000	0.000	0.000
60	15	0.000	0.003	0.003	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero.

Table F2

Results from the Simulation Study for Raw Intraclass Correlations (ICCs) Aggregated Across Replications when Setting Negative ICCs to Zero (Negative Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	5	.278	.563	.772	-0.403	-0.052	0.175	0.064	0.143	0.485	0.123	0.168	0.213	.000	.403	.667	.009	.098	.211
14	10	.409	.650	.839	-0.403	-0.027	0.100	0.043	0.107	0.483	0.118	0.150	0.183	.351	.574	.724	.000	.021	.083
14	15	.491	.686	.860	-0.394	-0.019	0.051	0.034	0.095	0.475	0.116	0.144	0.174	.417	.627	.741	.000	.008	.046
20	5	.354	.633	.829	-0.364	-0.049	0.146	0.059	0.126	0.466	0.124	0.158	0.201	.283	.545	.737	.000	.069	.193
20	10	.509	.726	.879	-0.350	-0.021	0.065	0.037	0.089	0.454	0.110	0.141	0.170	.449	.670	.773	.000	.010	.055
20	15	.534	.759	.890	-0.342	-0.013	0.031	0.026	0.077	0.446	0.114	0.136	0.159	.583	.708	.795	.000	.003	.028
30	5	.440	.716	.874	-0.276	-0.045	0.116	0.056	0.108	0.417	0.115	0.148	0.189	.448	.663	.811	.000	.043	.119
30	10	.627	.809	.927	-0.265	-0.016	0.043	0.029	0.070	0.402	0.109	0.132	0.158	.649	.753	.826	.000	.003	.028
30	15	.691	.843	.933	-0.251	-0.008	0.017	0.019	0.058	0.390	0.108	0.127	0.145	.703	.779	.831	.000	.000	.018
40	5	.519	.774	.900	-0.208	-0.043	0.090	0.054	0.096	0.370	0.115	0.141	0.173	.604	.728	.835	.000	.030	.110
40	10	.716	.868	.946	-0.194	-0.014	0.027	0.025	0.057	0.349	0.103	0.127	0.149	.693	.798	.853	.000	.001	.018
40	15	.765	.901	.961	-0.179	-0.005	0.008	0.014	0.044	0.334	0.107	0.122	0.137	.766	.820	.860	.000	.000	.009
50	5	.627	.819	.922	-0.131	-0.041	0.071	0.052	0.087	0.304	0.113	0.137	0.167	.675	.771	.859	.000	.022	.083
50	10	.750	.912	.966	-0.128	-0.011	0.018	0.021	0.047	0.287	0.104	0.123	0.139	.762	.829	.872	.000	.000	.018
50	15	.820	.943	.985	-0.124	-0.003	0.006	0.011	0.033	0.280	0.104	0.119	0.131	.804	.846	.878	.000	.000	.000
60	5	.650	.854	.928	-0.103	-0.040	0.059	0.047	0.080	0.237	0.110	0.134	0.161	.718	.801	.876	.000	.018	.073
60	10	.798	.946	.980	-0.061	-0.010	0.013	0.018	0.038	0.201	0.105	0.121	0.133	.808	.851	.879	.000	.000	.009
60	15	.863	.978	.994	-0.054	-0.001	0.003	0.007	0.021	0.190	0.107	0.117	0.124	.842	.866	.887	.000	.000	.000
70	15	-	-	-	-	-	-	-	-	-	-	0.115	-	-	.881	-	-	.000	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 15 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The number of participants that were used for the reliability analysis was 109 across all conditions and replications. There were no upper bound (ICCs ≥ 1) or lower bound (ICCs ≤ -1/[k-1]) estimation problems in any condition. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criteria, they reflect the minimum and maximum value across replications. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table F3

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Setting Negative ICCs to Zero Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Positive Emotion Differentiation in the Primary Data Set)

Condition		MCSE					
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC
14	4	0.001	0.003	0.002	0.001	0.000	0.000
14	8	0.001	0.002	0.002	0.001	0.000	0.000
14	12	0.001	0.002	0.002	0.001	0.000	0.000
20	4	0.001	0.003	0.002	0.001	0.000	0.000
20	8	0.000	0.002	0.002	0.001	0.000	0.000
20	12	0.000	0.002	0.002	0.001	0.000	0.000
30	4	0.001	0.003	0.002	0.001	0.000	0.000
30	8	0.000	0.002	0.001	0.000	0.000	0.000
30	12	0.000	0.001	0.001	0.000	0.000	0.000
40	4	0.000	0.003	0.002	0.001	0.000	0.000
40	8	0.000	0.002	0.001	0.000	0.000	0.000
40	12	0.000	0.001	0.001	0.000	0.000	0.000
50	4	0.000	0.002	0.002	0.000	0.000	0.000
50	8	0.000	0.001	0.001	0.000	0.000	0.000
50	12	0.000	0.001	0.001	0.000	0.000	0.000
60	4	0.000	0.002	0.002	0.000	0.000	0.000
60	8	0.000	0.001	0.001	0.000	0.000	0.000
60	12	0.000	0.000	0.001	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero.

Table F4

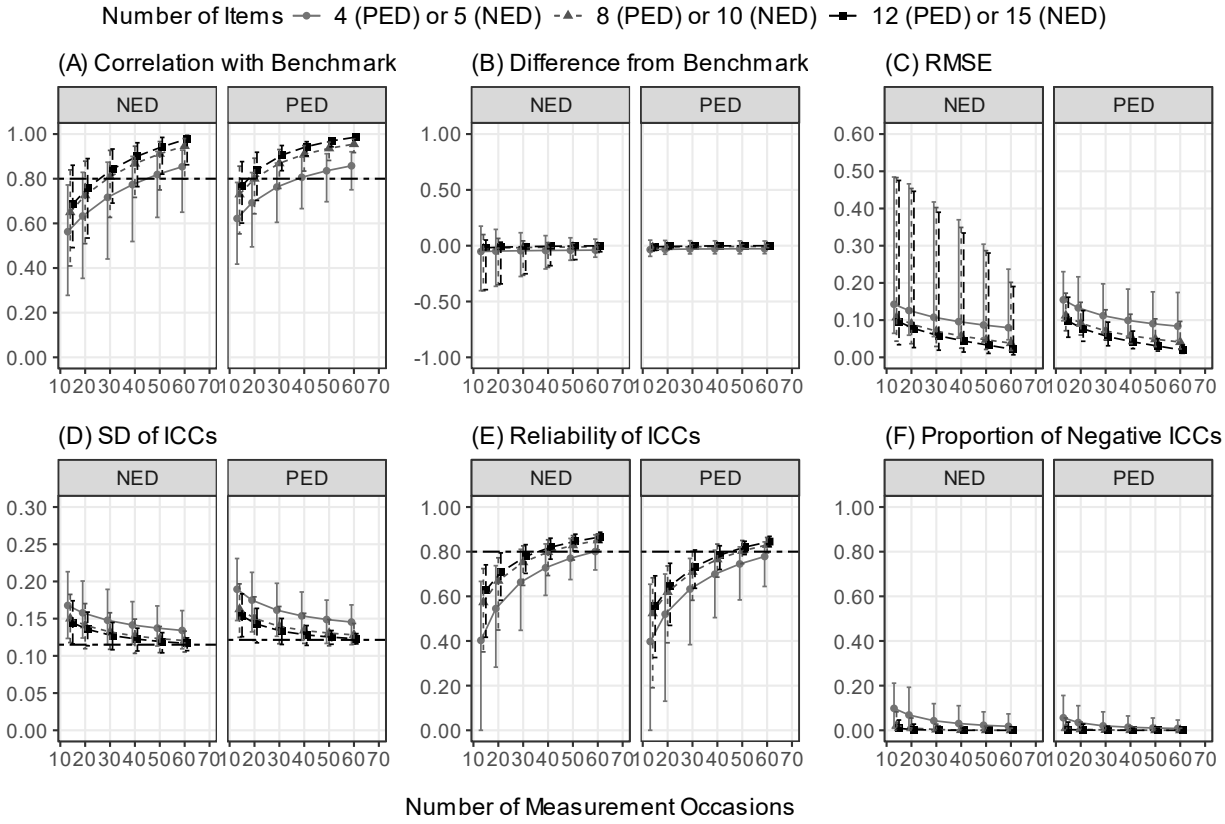
Results from the Simulation Study for Raw Intraclass Correlations (ICCs) Aggregated Across Replications when Setting Negative ICCs to Zero (Positive Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	4	.418	.622	.783	-0.095	-0.035	0.050	0.098	0.155	0.230	0.147	0.189	0.231	.000	.397	.654	.000	.056	.156
14	8	.553	.730	.856	-0.049	-0.016	0.014	0.071	0.112	0.172	0.125	0.163	0.197	.191	.524	.683	.000	.010	.055
14	12	.601	.769	.877	-0.042	-0.010	0.010	0.054	0.097	0.162	0.126	0.154	0.183	.326	.557	.692	.000	.003	.037
20	4	.495	.692	.828	-0.076	-0.032	0.047	0.084	0.133	0.216	0.132	0.175	0.212	.130	.520	.700	.000	.035	.110
20	8	.643	.800	.892	-0.037	-0.012	0.009	0.060	0.092	0.148	0.121	0.150	0.179	.392	.620	.736	.000	.005	.037
20	12	.703	.839	.919	-0.028	-0.006	0.008	0.043	0.076	0.126	0.118	0.143	0.164	.470	.647	.748	.000	.001	.028
30	4	.605	.763	.870	-0.073	-0.030	0.045	0.068	0.112	0.197	0.126	0.161	0.198	.384	.633	.770	.000	.020	.083
30	8	.756	.869	.931	-0.025	-0.009	0.008	0.047	0.071	0.128	0.116	0.140	0.166	.582	.709	.801	.000	.002	.018
30	12	.836	.906	.949	-0.016	-0.003	0.004	0.031	0.055	0.095	0.115	0.133	0.150	.636	.733	.807	.000	.000	.018
40	4	.666	.806	.897	-0.072	-0.029	0.044	0.057	0.099	0.183	0.121	0.154	0.186	.505	.699	.807	.000	.014	.064
40	8	.834	.909	.948	-0.018	-0.008	0.009	0.037	0.058	0.116	0.117	0.134	0.157	.686	.766	.834	.000	.001	.018
40	12	.902	.944	.966	-0.010	-0.002	0.002	0.023	0.042	0.071	0.114	0.128	0.141	.720	.785	.826	.000	.000	.009
50	4	.697	.836	.912	-0.073	-0.028	0.043	0.050	0.090	0.176	0.117	0.149	0.175	.584	.745	.822	.000	.010	.055
50	8	.882	.936	.964	-0.017	-0.007	0.009	0.029	0.049	0.103	0.113	0.131	0.149	.746	.803	.850	.000	.000	.009
50	12	.943	.969	.983	-0.005	-0.001	0.001	0.017	0.030	0.050	0.116	0.125	0.134	.787	.820	.847	.000	.000	.009
60	4	.75	.858	.921	-0.071	-0.027	0.042	0.045	0.084	0.174	0.115	0.145	0.168	.644	.778	.844	.000	.008	.046
60	8	.917	.955	.977	-0.017	-0.006	0.009	0.022	0.041	0.097	0.116	0.128	0.143	.795	.830	.864	.000	.000	.009
60	12	.974	.987	.994	-0.003	0.000	0.001	0.011	0.020	0.033	0.116	0.123	0.130	.824	.845	.869	.000	.000	.000
70	12	-	-	-	-	-	-	-	-	-	-	0.121	-	-	.864	-	-	.000	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 12 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The number of participants that were used for the reliability analysis was 109 across all conditions and replications. There were no upper bound (ICCs ≥ 1) or lower bound (ICCs ≤ -1/[k-1]) estimation problems in any condition. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criteria, they reflect the minimum and maximum value across replications. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Figure F1

Results of the Evaluation Criteria for Raw Intraclass Correlations (ICCs) for Each Condition When Setting Negative ICCs to Zero



Note. Panels (A) to (F) depict six evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (F) are absolute criteria. The panels present results for negative emotion differentiation (NED; left) and positive emotion differentiation (PED; right). The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across 5,000 replications. The plots display mean values for each item number condition. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E), and the standard deviation (*SD*) of ICCs in the benchmark condition in (D). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs.

Table F5

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Excluding Negative ICCs Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Negative Emotion Differentiation in the Primary Data Set)

Condition		MCSE					
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC
14	5	0.001	0.004	0.003	0.002	0.000	0.000
14	10	0.001	0.004	0.002	0.001	0.000	0.000
14	15	0.001	0.004	0.002	0.001	0.000	0.000
20	5	0.001	0.005	0.002	0.001	0.000	0.000
20	10	0.001	0.004	0.002	0.001	0.000	0.000
20	15	0.001	0.004	0.002	0.000	0.000	0.000
30	5	0.001	0.005	0.003	0.001	0.000	0.000
30	10	0.001	0.004	0.002	0.000	0.000	0.000
30	15	0.001	0.004	0.002	0.000	0.000	0.000
40	5	0.001	0.004	0.003	0.001	0.000	0.000
40	10	0.001	0.004	0.003	0.000	0.000	0.000
40	15	0.001	0.004	0.003	0.000	0.000	0.000
50	5	0.001	0.004	0.003	0.001	0.000	0.000
50	10	0.001	0.004	0.003	0.000	0.000	0.000
50	15	0.001	0.004	0.003	0.000	0.000	0.000
60	5	0.001	0.003	0.003	0.000	0.000	0.000
60	10	0.000	0.003	0.003	0.000	0.000	0.000
60	15	0.000	0.003	0.003	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero.

Table F6

Results from the Simulation Study for Raw Intraclass Correlations (ICCs) Aggregated Across Replications when Excluding Negative ICCs (Negative Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	5	.234	.546	.777	-0.359	-0.037	0.183	0.055	0.135	0.453	0.116	0.161	0.213	.000	.325	.642	.009	.099	.220
14	10	.435	.652	.825	-0.381	-0.024	0.098	0.041	0.106	0.467	0.116	0.148	0.180	.297	.542	.722	.000	.021	.073
14	15	.483	.686	.868	-0.394	-0.018	0.049	0.033	0.095	0.472	0.116	0.143	0.169	.441	.611	.728	.000	.008	.046
20	5	.306	.622	.829	-0.313	-0.039	0.150	0.048	0.120	0.436	0.116	0.152	0.197	.135	.494	.732	.000	.068	.174
20	10	.501	.724	.885	-0.343	-0.020	0.065	0.034	0.089	0.448	0.112	0.139	0.171	.475	.655	.776	.000	.011	.064
20	15	.541	.761	.901	-0.331	-0.013	0.032	0.027	0.077	0.440	0.112	0.135	0.161	.559	.703	.796	.000	.003	.028
30	5	.466	.706	.873	-0.254	-0.040	0.110	0.044	0.104	0.400	0.109	0.143	0.182	.416	.633	.802	.000	.043	.119
30	10	.622	.809	.931	-0.263	-0.016	0.042	0.029	0.070	0.400	0.110	0.131	0.159	.630	.748	.822	.000	.003	.037
30	15	.700	.842	.940	-0.262	-0.008	0.017	0.019	0.058	0.394	0.108	0.127	0.145	.698	.778	.834	.000	.000	.018
40	5	.500	.767	.904	-0.186	-0.040	0.088	0.042	0.094	0.353	0.106	0.138	0.173	.519	.710	.834	.000	.030	.101
40	10	.707	.868	.947	-0.190	-0.013	0.028	0.024	0.057	0.346	0.108	0.127	0.151	.725	.797	.856	.000	.001	.018
40	15	.772	.900	.964	-0.186	-0.005	0.008	0.015	0.044	0.339	0.107	0.122	0.137	.764	.819	.861	.000	.000	.009
50	5	.565	.811	.925	-0.137	-0.039	0.071	0.039	0.085	0.307	0.110	0.135	0.163	.640	.757	.855	.000	.023	.083
50	10	.761	.912	.963	-0.125	-0.011	0.018	0.021	0.047	0.284	0.106	0.123	0.140	.771	.828	.871	.000	.000	.018
50	15	.814	.944	.983	-0.123	-0.003	0.004	0.010	0.033	0.279	0.107	0.119	0.130	.811	.847	.875	.000	.000	.000
60	5	.632	.850	.927	-0.099	-0.038	0.057	0.036	0.079	0.232	0.106	0.132	0.159	.693	.791	.87	.000	.017	.073
60	10	.785	.946	.977	-0.070	-0.010	0.013	0.018	0.038	0.212	0.106	0.120	0.134	.811	.851	.876	.000	.000	.009
60	15	.863	.978	.994	-0.058	-0.001	0.001	0.007	0.021	0.194	0.105	0.117	0.125	.841	.866	.886	.000	.000	.000
70	15	-	-	-	-	-	-	-	-	-	-	0.115	-	-	.881	-	-	.000	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 15 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The number of participants that were used for the reliability analysis was 109 across all conditions and replications. There were no upper bound (ICCs $\geq$ 1) or lower bound (ICCs $\leq$ -1/[k-1]) estimation problems in any condition. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criteria, they reflect the minimum and maximum value across replications. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table F7

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Excluding Negative ICCs Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Positive Emotion Differentiation in the Primary Data Set)

Condition		MCSE					
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC
14	4	0.001	0.003	0.003	0.002	0.002	0.000
14	8	0.001	0.002	0.002	0.001	0.001	0.000
14	12	0.001	0.002	0.002	0.001	0.001	0.000
20	4	0.001	0.003	0.003	0.001	0.001	0.000
20	8	0.000	0.002	0.002	0.001	0.001	0.000
20	12	0.000	0.002	0.002	0.001	0.001	0.000
30	4	0.001	0.003	0.003	0.001	0.001	0.000
30	8	0.000	0.002	0.001	0.000	0.000	0.000
30	12	0.000	0.001	0.001	0.000	0.000	0.000
40	4	0.001	0.003	0.002	0.001	0.001	0.000
40	8	0.000	0.002	0.001	0.000	0.000	0.000
40	12	0.000	0.001	0.001	0.000	0.000	0.000
50	4	0.000	0.002	0.002	0.001	0.001	0.000
50	8	0.000	0.001	0.001	0.000	0.000	0.000
50	12	0.000	0.001	0.001	0.000	0.000	0.000
60	4	0.000	0.002	0.002	0.000	0.000	0.000
60	8	0.000	0.001	0.001	0.000	0.000	0.000
60	12	0.000	0.000	0.000	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero.

Table F8

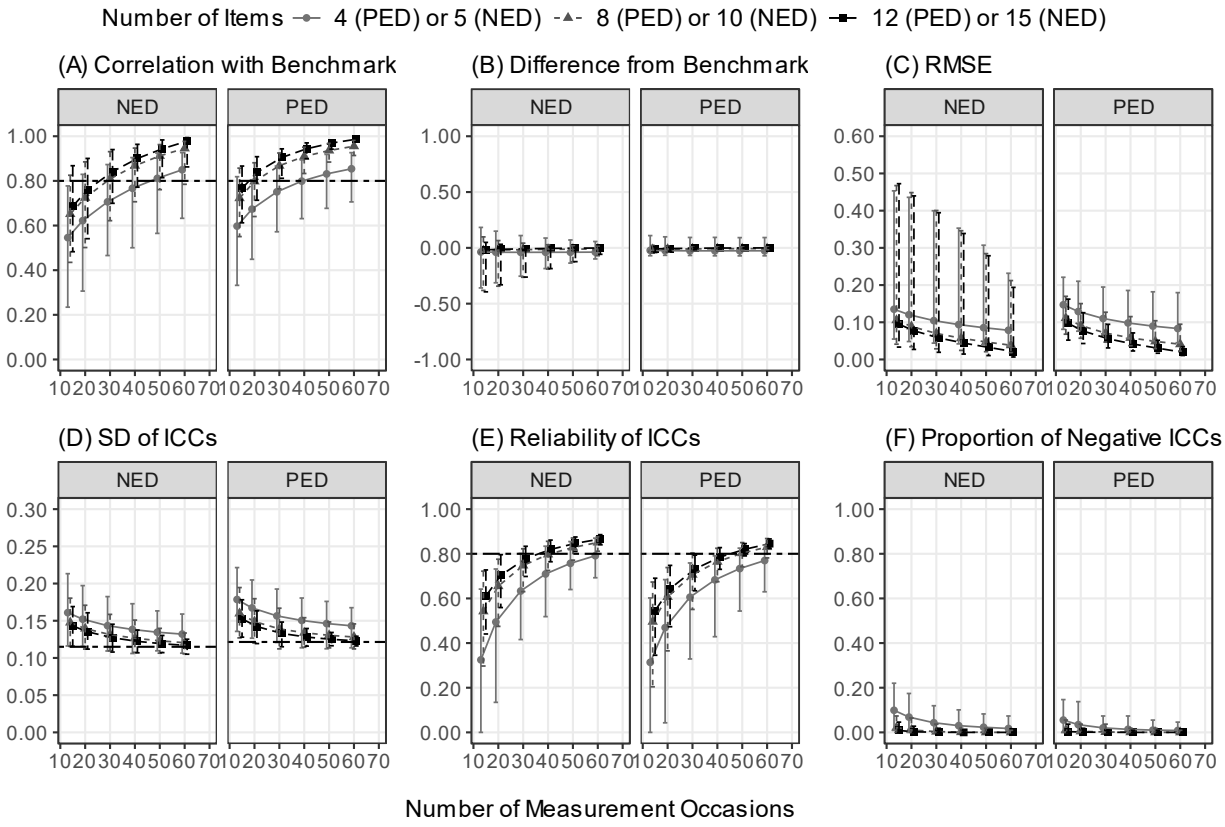
Results from the Simulation Study for Raw Intraclass Correlations (ICCs) Aggregated Across Replications when Excluding Negative ICCs (Positive Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	4	.332	.597	.819	-0.071	-0.022	0.110	0.081	0.147	0.221	0.136	0.179	0.221	.000	.314	.602	.000	.055	.147
14	8	.550	.723	.857	-0.047	-0.014	0.022	0.068	0.111	0.170	0.127	0.160	0.195	.204	.496	.673	.000	.010	.055
14	12	.612	.766	.866	-0.041	-0.010	0.014	0.052	0.097	0.162	0.128	0.153	0.178	.346	.546	.691	.000	.003	.037
20	4	.449	.674	.805	-0.067	-0.025	0.099	0.076	0.129	0.210	0.127	0.167	0.205	.043	.469	.684	.000	.034	.138
20	8	.640	.798	.881	-0.036	-0.012	0.015	0.057	0.091	0.150	0.120	0.149	0.180	.365	.607	.738	.000	.004	.037
20	12	.713	.838	.908	-0.030	-0.006	0.006	0.043	0.076	0.126	0.119	0.142	0.163	.473	.644	.748	.000	.001	.018
30	4	.572	.751	.866	-0.068	-0.027	0.093	0.069	0.110	0.195	0.121	0.157	0.193	.329	.606	.759	.000	.020	.073
30	8	.764	.868	.925	-0.025	-0.009	0.009	0.047	0.071	0.127	0.112	0.139	0.167	.551	.706	.803	.000	.002	.037
30	12	.812	.905	.944	-0.017	-0.003	0.005	0.031	0.056	0.094	0.115	0.133	0.148	.634	.732	.794	.000	.000	.009
40	4	.631	.799	.894	-0.071	-0.027	0.094	0.057	0.098	0.185	0.114	0.150	0.181	.429	.683	.791	.000	.014	.073
40	8	.833	.909	.951	-0.018	-0.008	0.009	0.037	0.058	0.115	0.114	0.134	0.154	.675	.765	.825	.000	.001	.018
40	12	.899	.944	.970	-0.011	-0.002	0.002	0.023	0.042	0.071	0.116	0.128	0.140	.730	.785	.828	.000	.000	.009
50	4	.677	.831	.919	-0.072	-0.026	0.092	0.050	0.009	0.182	0.112	0.146	0.176	.544	.734	.825	.000	.010	.055
50	8	.884	.936	.968	-0.018	-0.007	0.009	0.029	0.049	0.104	0.114	0.130	0.148	.742	.803	.850	.000	.000	.018
50	12	.942	.969	.983	-0.005	-0.001	0.001	0.017	0.030	0.051	0.116	0.125	0.134	.791	.820	.846	.000	.000	.009
60	4	.706	.854	.927	-0.071	-0.026	0.092	0.045	0.083	0.180	0.114	0.143	0.168	.630	.770	.839	.000	.009	.046
60	8	.914	.955	.981	-0.017	-0.006	0.008	0.022	0.041	0.095	0.112	0.128	0.146	.781	.830	.868	.000	.000	.009
60	12	.974	.987	.993	-0.002	0.000	0.001	0.011	0.020	0.034	0.116	0.123	0.129	.825	.845	.867	.000	.000	.000
70	12	-	-	-	-	-	-	-	-	-	-	0.121	-	-	.864	-	-	.000	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 15 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The number of participants that were used for the reliability analysis was 109 across all conditions and replications. There were no upper bound (ICCs ≥ 1) or lower bound (ICCs ≤ -1/[k-1]) estimation problems in any condition. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criteria, they reflect the minimum and maximum value across replications. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Figure F2

Results of the Evaluation Criteria for raw Intraclass Correlations (ICCs) for Each Condition When Excluding Negative ICCs



Note. Panels (A) to (F) depict six evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (F) are absolute criteria. The panels present results for negative emotion differentiation (NED; left) and positive emotion differentiation (PED; right). The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across 5,000 replications. The plots display mean values for each item number condition. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E), and the standard deviation (*SD*) of ICCs in the benchmark condition in (D). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs.

G: Sensitivity Analysis Comparing Item Sets with Lowest and Highest Endorsement

To compare item sets with the lowest and the highest endorsement, we identified the item sets with the highest and the lowest mean value in the benchmark sample.

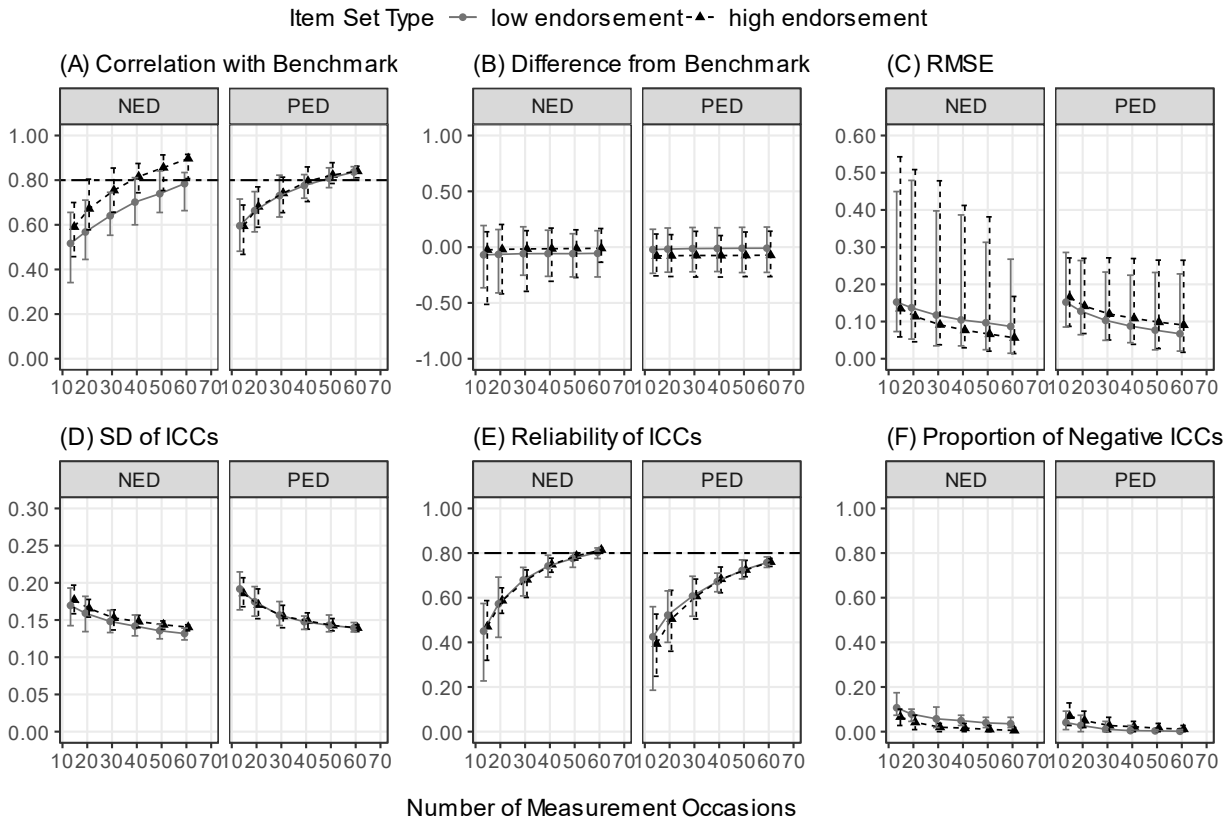
The sample means of the full item sets were 16.02 for negative emotion differentiation (15 items) and 44.77 for positive emotion differentiation (12 items).

For negative emotion differentiation, the 5-item set with the lowest mean value consisted of angry, sad, terrified, disgraced, and guilty (M = 12.80), whereas the 5-item set with the highest mean value consisted of annoyed, unhappy, worried, ashamed, and repentant (M = 19.67). The 10-item set with the lowest mean value consisted of angry, irritated, sad, downhearted, terrified, fearful, humiliated, disgraced, guilty, and 'I had a bad conscience' (M = 14.19), and the 10-item set with the highest mean value consisted of irritated, annoyed, downhearted, unhappy, fearful, worried, ashamed, humiliated, repentant, and 'I had a bad conscience' (M = 17.62).

For positive emotion differentiation, the 4-item set with the lowest mean value consisted of joyful, curious, trust, proud (M = 39.68), whereas the 4-item set with the highest mean value included happy, alert, closeness, confident (M = 49.26). The 8-item set with the lowest mean value included joyful, glad, interested, curious, love, trust, proud, self-assured (M = 42.52), and the 8-item set with the highest mean value included glad, happy, interested, alert, love, closeness, confident, self-assured (M = 47.32).

Figure G1

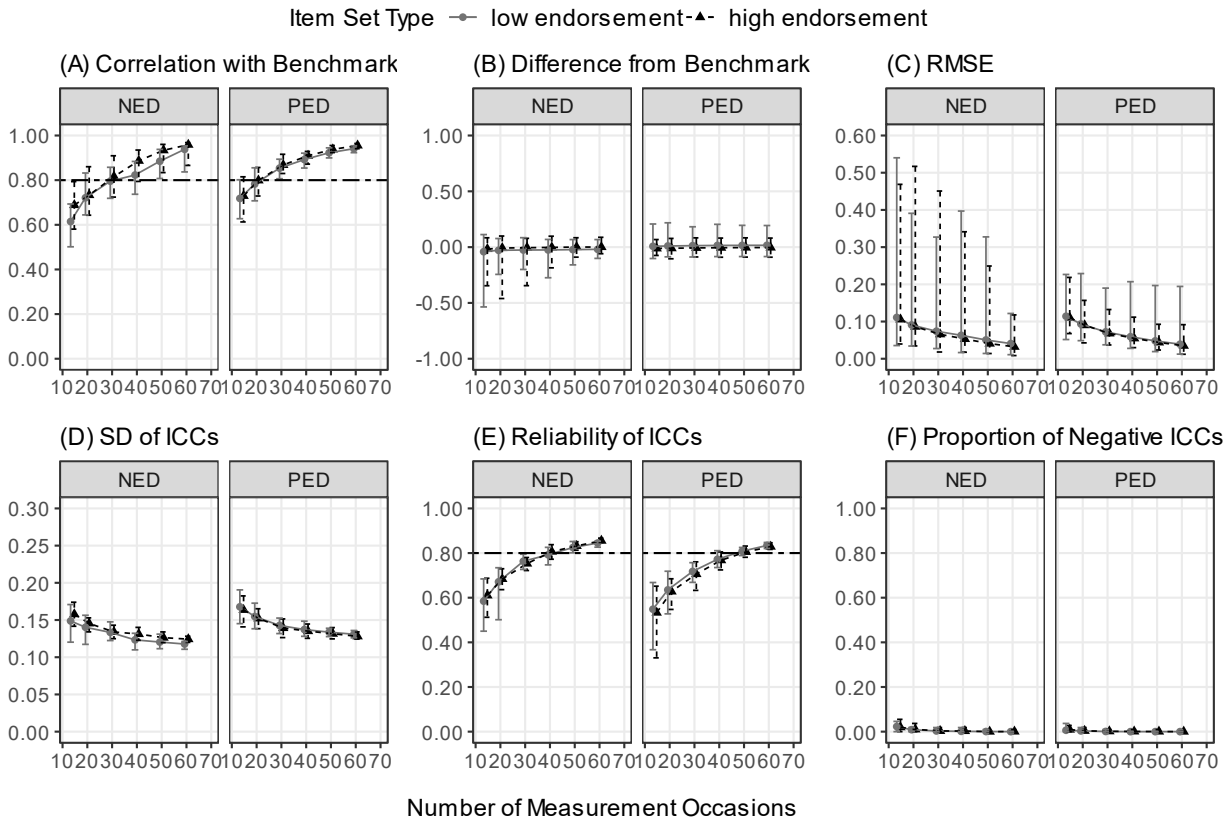
Results of the Evaluation Criteria for raw Intraclass Correlations (ICCs) Separate for the Item Sets with the Highest and the Lowest Endorsement in the Sample (4 and 5 Items)



Note. Panels (A) to (F) depict six evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (F) are absolute criteria. The panels present results for 5 items for negative emotion differentiation (NED; left) and 4 items for positive emotion differentiation (PED; right) in dependence of the item endorsement. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 20 to 21 replications (NED) or 61 to 62 replications (PED), respectively, per occasion-number condition. The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across replications. The plots display mean values for low and high endorsement item sets. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs.

Figure G2

Results of the Evaluation Criteria for raw Intraclass Correlations (ICCs) Separate for the Item Sets with the Highest and the Lowest Endorsement in the Sample (8 and 10 Items)



Note. Panels (A) to (F) depict six evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (F) are absolute criteria. The panels present results for 10 items for negative emotion differentiation (NED; left) and 8 items for positive emotion differentiation (PED; right) in dependence of the item endorsement. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 20 to 21 replications (NED) or 61 to 62 replications (PED), respectively, per occasion-number condition. The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across replications. The plots display mean values for low and high endorsement item sets. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs.

Table G1

Simulation Study Results When Only Comparing the Item Sets with the Lowest and the Highest Endorsement (Negative Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	End.	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 5 items																			
14	low	.341	.517	.656	-0.365	-0.069	0.193	0.073	0.152	0.449	0.142	0.170	0.193	.227	.450	.574	.073	.107	.174
20	low	.445	.568	.711	-0.410	-0.065	0.158	0.053	0.137	0.479	0.134	0.159	0.182	.423	.572	.692	.046	.078	.101
30	low	.553	.640	.746	-0.252	-0.059	0.182	0.035	0.117	0.397	0.133	0.148	0.163	.608	.679	.736	.018	.057	.110
40	low	.600	.702	.811	-0.262	-0.058	0.151	0.035	0.104	0.387	0.129	0.142	0.157	.692	.742	.790	.028	.049	.073
50	low	.655	.739	.841	-0.267	-0.059	0.120	0.024	0.097	0.313	0.125	0.136	0.144	.736	.776	.800	.018	.039	.064
60	low	.664	.784	.834	-0.267	-0.056	0.147	0.015	0.087	0.268	0.123	0.132	0.139	.776	.803	.823	.018	.035	.064
14	high	.457	.591	.700	-0.513	-0.025	0.137	0.059	0.135	0.543	0.158	0.177	0.197	.320	.471	.587	.028	.066	.101
20	high	.578	.672	.804	-0.418	-0.020	0.203	0.046	0.114	0.508	0.154	0.166	0.178	.530	.587	.644	.009	.042	.073
30	high	.656	.754	.854	-0.396	-0.016	0.147	0.037	0.091	0.478	0.137	0.153	0.164	.601	.681	.724	.000	.020	.037
40	high	.744	.816	.875	-0.306	-0.014	0.170	0.029	0.077	0.412	0.140	0.148	0.156	.714	.749	.777	.000	.016	.037
50	high	.754	.856	.913	-0.272	-0.012	0.155	0.020	0.066	0.381	0.137	0.144	0.149	.769	.787	.801	.000	.010	.028
60	high	.807	.897	.915	-0.136	-0.011	0.167	0.014	0.056	0.167	0.136	0.140	0.144	.801	.814	.824	.000	.005	.018

Table G1 (continued)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	End.	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 10 items																			
14	low	.502	.615	.693	-0.536	-0.041	0.112	0.035	0.111	0.540	0.120	0.149	0.171	.450	.585	.684	.000	.023	.046
20	low	.644	.721	.832	-0.244	-0.029	0.077	0.034	0.091	0.391	0.117	0.140	0.156	.501	.671	.734	.000	.009	.018
30	low	.719	.795	.858	-0.200	-0.026	0.084	0.027	0.074	0.327	0.122	0.133	0.148	.727	.763	.795	.000	.003	.018
40	low	.737	.823	.884	-0.274	-0.025	0.069	0.017	0.063	0.397	0.110	0.123	0.132	.747	.792	.826	.000	.001	.018
50	low	.807	.885	.938	-0.159	-0.022	0.068	0.014	0.051	0.328	0.111	0.120	0.131	.801	.825	.852	.000	.000	.009
60	low	.837	.938	.954	-0.101	-0.02	0.068	0.011	0.041	0.122	0.111	0.118	0.121	.826	.845	.853	.000	.000	.000
14	high	.581	.688	.794	-0.346	-0.017	0.084	0.040	0.106	0.469	0.142	0.158	0.174	.512	.610	.687	.000	.020	.055
20	high	.643	.733	.860	-0.460	-0.008	0.099	0.034	0.086	0.517	0.134	0.145	0.153	.636	.684	.730	.000	.010	.037
30	high	.724	.816	.909	-0.345	-0.005	0.079	0.018	0.066	0.451	0.125	0.134	0.143	.722	.753	.780	.000	.003	.018
40	high	.794	.887	.935	-0.186	-0.002	0.098	0.018	0.053	0.341	0.123	0.131	0.140	.772	.806	.838	.000	.003	.018
50	high	.834	.934	.960	-0.090	0.001	0.084	0.014	0.041	0.249	0.120	0.126	0.134	.813	.832	.851	.000	.000	.000
60	high	.866	.959	.968	-0.059	0.001	0.088	0.009	0.032	0.118	0.120	0.124	0.128	.848	.855	.867	.000	.000	.009

Note. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 20 to 21 replications per occasion-number condition. The mean values were *M* = 12.80 for the low endorsement 5-item set, *M* = 19.67 for the high endorsement 5-item set, *M* = 14.19 for the low endorsement 10-item set, and *M* = 17.62 for the high endorsement 10-item set. Since the 15-item conditions included the full item set, we could not compare low and high endorsement item sets for this number of items. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. End. = endorsement (sample mean value of item set). Min = minimum. *M* = mean. Max = maximum. End. = endorsement (sample mean value of item set).

Table G2

Simulation Study Results When Only Comparing the Item Sets with the Lowest and the Highest Endorsement (Positive Emotion Differentiation in the Primary Data Set)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	End.	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 4 items																			
14	low	.482	.596	.715	-0.235	-0.021	0.160	0.085	0.152	0.286	0.164	0.192	0.215	.185	.425	.560	.009	.041	.092
20	low	.568	.663	.749	-0.224	-0.018	0.171	0.065	0.128	0.264	0.155	0.175	0.195	.400	.522	.630	.000	.028	.073
30	low	.635	.731	.822	-0.221	-0.013	0.176	0.049	0.103	0.233	0.143	0.157	0.175	.517	.607	.696	.000	.011	.046
40	low	.719	.776	.825	-0.219	-0.012	0.174	0.043	0.088	0.225	0.137	0.148	0.155	.627	.672	.710	.000	.004	.028
50	low	.766	.809	.855	-0.228	-0.010	0.178	0.024	0.077	0.232	0.134	0.143	0.157	.684	.722	.770	.000	.003	.009
60	low	.802	.836	.860	-0.226	-0.010	0.180	0.021	0.067	0.228	0.134	0.139	0.147	.736	.757	.783	.000	.002	.009
14	high	.467	.594	.688	-0.257	-0.079	0.117	0.087	0.165	0.271	0.168	0.186	0.207	.248	.393	.527	.028	.070	.128
20	high	.589	.681	.770	-0.263	-0.077	0.112	0.068	0.141	0.270	0.152	0.170	0.192	.360	.505	.633	.018	.050	.092
30	high	.655	.741	.813	-0.268	-0.075	0.141	0.050	0.120	0.271	0.140	0.155	0.170	.505	.606	.683	.000	.027	.064
40	high	.704	.796	.859	-0.268	-0.075	0.104	0.039	0.109	0.269	0.138	0.149	0.160	.622	.684	.737	.000	.022	.046
50	high	.785	.823	.878	-0.264	-0.074	0.135	0.027	0.098	0.265	0.135	0.143	0.152	.694	.725	.767	.000	.015	.037
60	high	.811	.840	.862	-0.264	-0.072	0.143	0.017	0.091	0.265	0.134	0.139	0.144	.740	.760	.776	.000	.011	.028

Table G2 (continued)

Condition		Correlation			Difference			RMSE			*SD*			Reliability			% Neg ICC		
N_{occ}	End.	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 8 items																			
14	low	.627	.718	.802	-0.102	0.008	0.207	0.052	0.114	0.227	0.145	0.168	0.191	.367	.548	.668	.000	.006	.037
20	low	.708	.783	.855	-0.087	0.010	0.218	0.049	0.093	0.229	0.138	0.155	0.173	.528	.635	.719	.000	.003	.018
30	low	.806	.854	.894	-0.089	0.013	0.182	0.037	0.072	0.190	0.132	0.143	0.153	.669	.718	.757	.000	.001	.009
40	low	.854	.892	.921	-0.085	0.015	0.204	0.028	0.059	0.207	0.128	0.137	0.148	.735	.772	.811	.000	.000	.000
50	low	.900	.923	.945	-0.084	0.016	0.195	0.019	0.048	0.197	0.128	0.134	0.139	.789	.808	.825	.000	.000	.000
60	low	.923	.941	.954	-0.085	0.017	0.194	0.012	0.039	0.194	0.126	0.131	0.136	.820	.834	.847	.000	.000	.000
14	high	.613	.729	.815	-0.075	-0.012	0.068	0.068	0.110	0.219	0.141	0.164	0.182	.330	.533	.652	.000	.009	.028
20	high	.729	.799	.857	-0.104	-0.010	0.078	0.043	0.090	0.157	0.138	0.152	0.165	.547	.628	.685	.000	.003	.018
30	high	.830	.868	.916	-0.088	-0.007	0.083	0.037	0.068	0.132	0.126	0.139	0.151	.633	.706	.761	.000	.002	.009
40	high	.872	.910	.929	-0.090	-0.005	0.081	0.030	0.055	0.112	0.125	0.135	0.145	.725	.768	.805	.000	.001	.009
50	high	.923	.938	.955	-0.089	-0.004	0.083	0.023	0.044	0.093	0.125	0.132	0.140	.781	.805	.831	.000	.000	.009
60	high	.941	.955	.964	-0.090	-0.004	0.081	0.012	0.035	0.092	0.124	0.128	0.134	.816	.829	.846	.000	.000	.000

Note. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 61 to 62 replications per occasion-number condition. The mean values were $M = 39.68$ for the low endorsement 4-item set, $M = 49.26$ for the high endorsement 4-item set, $M = 42.52$ for the low endorsement 8-item set, and $M = 47.32$ for the high endorsement 8-item set. Since the 12-item conditions included the full item set, we could not compare low and high endorsement item sets for this number of items. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. End. = endorsement (sample mean value of item set). Min = minimum. *M* = mean. Max = maximum. End. = endorsement (sample mean value of item set).

H: Description of the Replication Data Set

Additional Information on Data Preparation

For the replication data set, we used existing data from the *EMOTIONS* project (Ryvkina et al., 2023). In total, 2,272 participants provided 64,810 occasions. Of these participants, 1,358 participated in Wave 1 only, 627 participated in Wave 2 only, and 287 participated in both waves. Because participants who took part in both waves completed the sociodemographic questionnaire twice, we retained the first available assessment of sociodemographic variables. Specifically, for participants who completed both waves, we used the sociodemographic information collected before Wave 1. For participants who completed only one wave, we used the sociodemographic information collected before their respective wave.

A total of 56 occasions were flagged for an outlying completion time (i.e., shorter than 1 s multiplied by the number of completed items; Ryvkina et al., 2023). Of these, 33 belonged to Wave 1 and 23 to Wave 2. After excluding these occasions, we prepared the emotion variables. At each occasion, participants first indicated whether they had experienced a social interaction of at least 5 minutes' duration since the previous occasion. If they had, they rated the subsequent items, including the emotion items, with respect to that social interaction. Otherwise, they rated the same items with respect to the last non-social activity of at least 5 minutes (Ryvkina et al., 2023). For the present research, we combined both situational contexts by including all occasions regardless of context. Because identical emotion items were assessed in both context but stored in separate variables, we collapsed the context-specific variables into a single variable for each emotion. For example, if a participant rated anger with respect to a social interaction, the non-social activity anger ratings was missing, and we retained the social interaction anger rating as the value for the unified anger variable.

Next, we combined data from Wave 1 and Wave 2 because of their temporal proximity. Specifically, occasions were ordered chronologically within each participant across

waves and assigned a consecutive occasion number. To create the benchmark data set, we then selected participants who provided at least 70 occasions in total. This resulted in a sample of 251 participants contributing 24,231 occasions. Of these participants, 150 participants completed both waves ($N_{occasions}$ = 16,653), 70 participated in Wave 1 only ($N_{occasions}$ = 5,262), and 31 participated in Wave 2 only ($N_{occasions}$ = 2,316). Then, we retained the first 70 occasions for each participant, resulting in 251 participants with 17,570 occasions.

Finally, because our analyses included all possible item sets for each number of emotion items, we checked whether each participant showed non-zero variance in emotion ratings across the 70 occasions included in the benchmark for every possible item set. Specifically, we determined all possible items sets with 3 to 9 items for negative emotion differentiation, and 3 to 5 items for positive emotion differentiation. For example, selecting 3 out of the 9 negative emotion items yields 84 unique item sets. We then calculated, separately for each participant, the variance in emotion ratings for each of these item sets across 70 occasions. In total, 75 participants showed zero variance for at least one negative emotion item sets (for more detailed information, see Table H1). No such cases occurred for the positive emotion item sets. These participants were excluded to ensure that all simulation conditions and replications were based on the same set of participants, thereby allowing fair comparisons across conditions. This approach was also consistent with the primary data set, in which no such variance-related problems occurred and all participants could therefore be included in every condition and replication. Thus, the final benchmark data set comprised 176 participants with 12,320 occasions. Of these participants, 105 participated in both waves, 50 participated in Wave 1 only, and 21 participated in Wave 2 only.

Table H1

Overview of Zero-Variance Cases Across Possible Negative Emotion Item Sets in the Replication Data Set

Item-number condition	Possible item sets	Affected item sets		Affected participants		Affected item sets per participant		
	N	*N*	%	*N*	%[a]	Min	Mdn	Max
3	84	84	100	75	30	1	4	84
4	126	126	100	42	17	1	5	126
5	126	126	100	25	10	1	1	126
6	84	84	100	10	4	1	17.5	84
7	36	36	100	8	3	1	8	36
8	9	9	100	5	2	1	1	9
9	1	1	100	1	0.4	1	1	1

Note. This table only refers to negative emotions since there were no cases with zero variance for positive emotions.

[a] The total number of participants with at least 70 occasions who were screened for zero variance in item sets was 251.

Table H2

Descriptive Statistics of Emotion Variables in the Replication Data Set

Variable	*M*	SD_{within}	$SD_{between}$	ICC	Range
Negative emotions					
angry	1.52	0.99	0.41	.15	1 – 6
socially excluded	1.53	0.82	0.77	.47	1 – 6
envious	1.17	0.50	0.29	.26	1 – 6
resentful	1.12	0.44	0.22	.21	1 – 6
ashamed	1.24	0.60	0.34	.25	1 – 6
insecure	1.71	0.94	0.63	.31	1 – 6
anxious	1.56	0.78	0.70	.44	1 – 6
sad	1.70	0.94	0.80	.43	1 – 6
lonely	1.77	0.88	0.99	.56	1 – 6
Positive emotions					
proud	2.30	1.15	0.88	.37	1 – 6
successful	2.78	1.28	0.97	.37	1 – 6
superior	1.56	0.81	0.66	.40	1 – 6
enthusiastic	2.52	1.28	0.75	.26	1 – 6
relaxed	3.78	1.14	0.91	.39	1 – 6

Note. *M* = grand mean. SD_{within} = within-person standard deviation. $SD_{between}$ = between-person standard deviation. ICC = intraclass correlation

Table H3

Overview of the Examined Conditions in the Replication Data Set

Number of occasions	Number of emotion items									
	NED							PED		
	3	4	5	6	7	8	9	3	4	5
14	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o
20	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o
30	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o
40	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o
50	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o
60	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o	r/o
70	o	o	o	o	o	o	benchmark	o	o	benchmark

Note. Each cell represents one condition resulting from crossing the number of occasions with the number of emotion items. Conditions marked with "r" were realized in the simulation using random draws of occasions (with R = 5,000 replications). Conditions marked with "o" were realized in the sensitivity analysis in which occasions were drawn in temporal order. Since drawing 70 occasions (i.e., the benchmark number) without replacement always reproduces the benchmark occasions, the 70-occasions conditions could not be meaningfully replicated for random draws of occasions. Thus, only temporally ordered draws of occasions apply in these conditions. Occasions were drawn per participant, and the same number was drawn for all participants within each condition. NED = negative emotion differentiation. PED = positive emotion differentiation.

Table H4

Total Number of Redraws Across Participants Separate for all Simulation Conditions in Replication of the Main Analysis (Replication Data Set)

Number of items	Number of occasions					
	14	20	30	40	50	60
			NED			
3	112,056	55,133	21,378	8,909	3,585	1,131
4	52,353	22,730	7,054	2,380	714	193
5	29,639	11,659	3,033	813	217	57
6	18,850	6,559	1,391	290	63	12
7	12,866	3,937	674	96	10	0
8	9,316	2,748	376	38	3	0
9	7,205	1,938	216	21	0	0
			PED			
3	1,287	517	143	44	19	1
4	167	13	0	0	0	0
5	28	2	0	0	0	0

Note. The simulation was based on drawing subsets of occasions and items. This practice may result in zero within-person variance for some participants, making intraclass correlation calculation impossible. Such cases would typically be excluded from analyses. However, to retain a sufficient number of participants across conditions, we implemented the re-drawing of occasions until each participant showed non-zero variance across items. This table refers to the replication of the main analysis with the replication data set. NED = negative emotion differentiation. PED = positive emotion differentiation.

J: Results for the Replication Data Set (Replication of Main Analysis)

Table J1

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Negative Emotion Differentiation in the Replication Data Set)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
14	3	0.001	0.006	0.004	0.002	0.000	0.001	0.000	0.022
14	4	0.001	0.006	0.004	0.001	0.000	0.001	0.000	0.014
14	5	0.001	0.006	0.004	0.001	0.000	0.001	0.000	0.011
14	6	0.001	0.006	0.003	0.003	0.000	0.000	0.000	0.004
14	7	0.001	0.006	0.003	0.001	0.000	0.000	0.000	0.009
14	8	0.001	0.006	0.003	0.002	0.000	0.000	0.000	0.003
14	9	0.001	0.006	0.003	0.001	0.000	0.000	0.000	0.008
20	3	0.001	0.006	0.004	0.001	0.000	0.001	0.000	0.020
20	4	0.001	0.006	0.004	0.001	0.000	0.001	0.000	0.013
20	5	0.001	0.006	0.004	0.001	0.000	0.001	0.000	0.010
20	6	0.001	0.006	0.003	0.001	0.000	0.000	0.000	0.008
20	7	0.001	0.006	0.003	0.000	0.000	0.000	0.000	0.007
20	8	0.001	0.006	0.003	0.000	0.000	0.000	0.000	0.006
20	9	0.001	0.006	0.003	0.000	0.000	0.000	0.000	0.006
30	3	0.001	0.006	0.004	0.002	0.000	0.001	0.000	0.016
30	4	0.001	0.005	0.004	0.002	0.000	0.001	0.000	0.006
30	5	0.001	0.005	0.004	0.000	0.000	0.001	0.000	0.008
30	6	0.001	0.005	0.004	0.001	0.000	0.000	0.000	0.005
30	7	0.001	0.005	0.004	0.001	0.000	0.000	0.000	0.003
30	8	0.001	0.005	0.004	0.001	0.000	0.000	0.000	0.002
30	9	0.001	0.005	0.004	0.000	0.000	0.000	0.000	0.002

Table J1 (continued)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
40	3	0.001	0.005	0.004	0.001	0.000	0.001	0.000	0.016
40	4	0.001	0.005	0.004	0.001	0.000	0.001	0.000	0.008
40	5	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.006
40	6	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.004
40	7	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.003
40	8	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.002
40	9	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.001
50	3	0.001	0.005	0.004	0.001	0.000	0.001	0.000	0.014
50	4	0.001	0.004	0.004	0.001	0.000	0.001	0.000	0.003
50	5	0.001	0.003	0.004	0.000	0.000	0.000	0.000	0.004
50	6	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.003
50	7	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.001
50	8	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.001
50	9	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.000
60	3	0.001	0.005	0.004	0.001	0.000	0.001	0.000	0.012
60	4	0.001	0.004	0.004	0.001	0.000	0.001	0.000	0.006
60	5	0.001	0.003	0.004	0.000	0.000	0.000	0.000	0.003
60	6	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.002
60	7	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.000
60	8	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.000
60	9	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero. LB estimation probl.= lower bound estimation problems (ICCs $\leq$ -1/[k-1]). UB estimation probl. = upper bound estimation problems (ICCs $\geq$ 1).

Table J2

Relative Evaluation Criteria for Raw Intraclass Correlations (ICCs) Aggregated Across Replications (Negative Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	.208	.493	.709	-0.321	-0.054	0.158	0.084	0.214	0.542
14	4	.240	.551	.783	-0.353	-0.047	0.069	0.068	0.179	0.548
14	5	.320	.588	.781	-0.383	-0.042	0.020	0.055	0.157	0.560
14	6	.402	.617	.793	-0.387	-0.039	0.014	0.045	0.142	0.559
14	7	.392	.634	.811	-0.402	-0.036	0.011	0.036	0.131	0.565
14	8	.425	.651	.825	-0.400	-0.034	0.007	0.029	0.122	0.562
14	9	.450	.663	.848	-0.396	-0.032	0.007	0.023	0.115	0.558
20	3	.280	.540	.733	-0.252	-0.045	0.157	0.083	0.197	0.489
20	4	.332	.606	.784	-0.286	-0.037	0.065	0.063	0.161	0.498
20	5	.367	.650	.810	-0.297	-0.032	0.018	0.049	0.139	0.501
20	6	.441	.683	.828	-0.296	-0.029	0.012	0.039	0.124	0.497
20	7	.475	.705	.844	-0.299	-0.026	0.009	0.030	0.112	0.498
20	8	.523	.723	.860	-0.309	-0.024	0.006	0.023	0.103	0.502
20	9	.525	.737	.884	-0.310	-0.022	0.005	0.017	0.095	0.501
30	3	.349	.603	.767	-0.171	-0.035	0.155	0.076	0.178	0.435
30	4	.382	.681	.832	-0.167	-0.027	0.081	0.058	0.141	0.397
30	5	.494	.731	.865	-0.179	-0.022	0.031	0.044	0.119	0.400
30	6	.552	.766	.881	-0.185	-0.019	0.012	0.032	0.103	0.401
30	7	.525	.791	.903	-0.201	-0.016	0.008	0.025	0.091	0.411
30	8	.579	.813	.922	-0.198	-0.015	0.004	0.018	0.081	0.407
30	9	.633	.831	.929	-0.193	-0.013	0.003	0.011	0.072	0.401

Table J2 (continued)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
40	3	.349	.650	.816	-0.132	-0.029	0.160	0.073	0.164	0.416
40	4	.497	.734	.859	-0.118	-0.022	0.075	0.055	0.128	0.328
40	5	.570	.790	.898	-0.106	-0.017	0.035	0.040	0.105	0.315
40	6	.622	.828	.918	-0.107	-0.013	0.013	0.030	0.088	0.308
40	7	.625	.857	.932	-0.104	-0.011	0.009	0.022	0.075	0.301
40	8	.683	.879	.942	-0.109	-0.009	0.004	0.016	0.065	0.304
40	9	.695	.896	.956	-0.109	-0.007	0.003	0.008	0.055	0.300
50	3	.438	.688	.821	-0.105	-0.025	0.152	0.072	0.154	0.396
50	4	.536	.776	.895	-0.082	-0.018	0.077	0.051	0.117	0.302
50	5	.628	.834	.917	-0.069	-0.013	0.035	0.038	0.094	0.228
50	6	.686	.874	.938	-0.060	-0.009	0.014	0.028	0.077	0.225
50	7	.743	.903	.954	-0.059	-0.007	0.009	0.021	0.063	0.214
50	8	.787	.926	.963	-0.056	-0.005	0.004	0.014	0.052	0.201
50	9	.797	.943	.978	-0.054	-0.004	0.002	0.006	0.040	0.205
60	3	.426	.716	.833	-0.085	-0.022	0.146	0.068	0.146	0.382
60	4	.623	.808	.910	-0.057	-0.015	0.075	0.050	0.109	0.279
60	5	.677	.868	.932	-0.040	-0.010	0.035	0.037	0.085	0.196
60	6	.769	.908	.953	-0.030	-0.007	0.015	0.027	0.068	0.145
60	7	.814	.937	.971	-0.028	-0.005	0.009	0.019	0.053	0.135
60	8	.841	.959	.981	-0.025	-0.003	0.005	0.012	0.040	0.124
60	9	.867	.977	.993	-0.022	-0.002	0.001	0.003	0.026	0.117

Note. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criteria, they reflect the minimum and maximum value across replications. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table J3

Absolute Evaluation Criteria for Raw Intraclass Correlations (ICCs) Aggregated Across Replications (Negative Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	0.175	0.238	0.312	.000	.430	.971	.091	.268	.500	0	0	0	0	1.698	11
14	4	0.160	0.211	0.270	.208	.520	.974	.085	.222	.449	0	0	0	0	1.010	6
14	5	0.150	0.196	0.244	.324	.572	.761	.057	.182	.381	0	0	1	0	0.721	5
14	6	0.127	0.186	0.231	.272	.757	.994	.028	.122	.273	0	0	0	0	0.086	2
14	7	0.141	0.180	0.221	.463	.638	.780	.034	.108	.205	0	0	0	0	0.451	3
14	8	0.140	0.175	0.209	.503	.752	.992	.023	.075	.165	0	0	0	0	0.054	2
14	9	0.139	0.171	0.206	.550	.677	.778	.011	.063	.125	0	0	0	0	0.315	4
20	3	0.167	0.226	0.302	.173	.580	.834	.068	.241	.477	0	0	0	0	1.382	11
20	4	0.150	0.201	0.253	.389	.646	.991	.040	.146	.324	0	0	0	0	0.798	6
20	5	0.147	0.186	0.232	.503	.686	.984	.040	.138	.324	0	0	0	0	0.522	5
20	6	0.139	0.178	0.213	.560	.713	.831	.017	.092	.233	0	0	0	0	0.367	4
20	7	0.134	0.172	0.208	.600	.733	.835	.017	.076	.176	0	0	0	0	0.299	3
20	8	0.133	0.167	0.201	.622	.748	.843	.006	.055	.119	0	0	0	0	0.230	2
20	9	0.131	0.163	0.193	.648	.760	.832	.006	.045	.097	0	0	0	0	0.175	3
30	3	0.147	0.213	0.295	.363	.753	.997	.040	.207	.432	0	0	0	0	0.835	9
30	4	0.150	0.189	0.245	.608	.819	.997	.011	.128	.312	0	0	0	0	0.159	3
30	5	0.134	0.176	0.218	.620	.773	.989	.017	.096	.278	0	0	0	0	0.312	4
30	6	0.139	0.168	0.204	.686	.802	.995	.006	.064	.170	0	0	0	0	0.128	3
30	7	0.134	0.162	0.191	.725	.818	.995	.006	.048	.136	0	0	0	0	0.046	2
30	8	0.131	0.158	0.183	.740	.825	.995	.000	.036	.080	0	0	0	0	0.029	1
30	9	0.127	0.155	0.182	.748	.830	.995	.000	.029	.062	0	0	0	0	0.021	1

Table J3 (continued)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
40	3	0.153	0.204	0.285	.558	.765	.991	.028	.182	.426	0	0	0	0	0.757	9
40	4	0.143	0.181	0.244	.669	.813	.996	.011	.110	.352	0	0	0	0	0.272	4
40	5	0.135	0.169	0.208	.728	.821	.992	.006	.073	.250	0	0	0	0	0.156	3
40	6	0.135	0.161	0.194	.752	.833	.992	.006	.048	.131	0	0	0	0	0.086	3
40	7	0.132	0.156	0.180	.785	.843	.905	.000	.034	.091	0	0	0	0	0.033	1
40	8	0.131	0.152	0.178	.803	.851	.898	.000	.026	.074	0	0	0	0	0.024	2
40	9	0.128	0.149	0.174	.806	.858	.993	.000	.021	.045	0	0	0	0	0.010	1
50	3	0.151	0.197	0.288	.614	.803	.992	.017	.164	.409	0	0	0	0	0.577	9
50	4	0.142	0.176	0.231	.729	.856	.998	.006	.091	.295	0	0	0	0	0.041	2
50	5	0.138	0.164	0.207	.780	.849	.994	.006	.058	.188	0	0	0	0	0.082	2
50	6	0.130	0.156	0.186	.807	.858	.926	.000	.037	.119	0	0	0	0	0.034	2
50	7	0.133	0.151	0.175	.820	.867	.994	.000	.026	.080	0	0	0	0	0.007	1
50	8	0.127	0.148	0.167	.834	.873	.907	.000	.019	.051	0	0	0	0	0.003	1
50	9	0.129	0.145	0.158	.848	.879	.907	.006	.016	.040	0	0	0	0	0.001	1
60	3	0.147	0.192	0.276	.671	.829	.994	.017	.150	.398	0	0	0	0	0.453	6
60	4	0.137	0.171	0.228	.768	.855	.997	.006	.080	.295	0	0	0	0	0.151	3
60	5	0.138	0.160	0.201	.818	.866	.995	.000	.047	.170	0	0	0	0	0.058	2
60	6	0.135	0.153	0.180	.834	.875	.995	.000	.030	.091	0	0	0	0	0.016	1
60	7	0.130	0.148	0.169	.849	.883	.995	.000	.020	.068	0	0	0	0	0.001	1
60	8	0.128	0.145	0.159	.858	.889	.915	.000	.015	.040	0	0	0	0	0	0
60	9	0.129	0.142	0.153	.875	.894	.911	.006	.014	.028	0	0	0	0	0	0
70	9	-	0.140	-	-	.906	-	-	.011	-	-	0	-	-	0	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 9 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The minimum and maximum reflect the minimum and maximum value across replications. The reliability calculation required Fisher's *Z*-transformed ICCs; thus, participants affected by estimation problems were excluded from these analyses. The minimum sample size for the reliability calculation ranged from 165 to 176 across conditions. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. LB = lower bound estimation problems (ICCs ≤ -1/[k-1]). UB = upper bound estimation problems (ICCs ≥ 1). N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table J4

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Positive Emotion Differentiation in the Replication Data Set)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
14	3	0.001	0.007	0.003	0.002	0.000	0.001	0.001	0.004
14	4	0.001	0.005	0.003	0.001	0.000	0.000	0.000	0.002
14	5	0.000	0.004	0.003	0.001	0.000	0.000	0.001	0.000
20	3	0.001	0.007	0.003	0.001	0.000	0.001	0.000	0.003
20	4	0.001	0.005	0.003	0.001	0.000	0.000	0.000	0.001
20	5	0.000	0.003	0.003	0.001	0.000	0.000	0.000	0.000
30	3	0.001	0.006	0.002	0.001	0.000	0.001	0.000	0.001
30	4	0.001	0.004	0.003	0.001	0.000	0.000	0.000	0.000
30	5	0.000	0.002	0.002	0.000	0.000	0.000	0.000	0.000
40	3	0.001	0.006	0.002	0.001	0.000	0.001	0.000	0.001
40	4	0.001	0.004	0.003	0.001	0.000	0.000	0.000	0.000
40	5	0.000	0.001	0.002	0.000	0.000	0.000	0.000	0.000
50	3	0.001	0.006	0.002	0.001	0.000	0.001	0.000	0.000
50	4	0.001	0.004	0.003	0.000	0.000	0.000	0.000	0.000
50	5	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000
60	3	0.001	0.006	0.002	0.001	0.000	0.001	0.000	0.000
60	4	0.001	0.004	0.003	0.000	0.000	0.000	0.000	0.000
60	5	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero. LB estimation probl.= lower bound estimation problems (ICCs $\leq$ -1/[k-1]). UB estimation probl. = upper bound estimation problems (ICCs $\geq$ 1).

Table J5

Relative Evaluation Criteria for Raw Intraclass Correlations (ICCs) Aggregated Across Replications (Positive Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	.338	.576	.764	-0.069	-0.022	0.023	0.140	0.200	0.462
14	4	.472	.667	.792	-0.091	-0.015	0.007	0.096	0.149	0.339
14	5	.577	.725	.831	-0.110	-0.011	0.010	0.060	0.116	0.276
20	3	.407	.641	.789	-0.053	-0.018	0.023	0.120	0.178	0.463
20	4	.521	.741	.844	-0.051	-0.011	0.013	0.078	0.127	0.319
20	5	.682	.804	.879	-0.064	-0.007	0.004	0.047	0.091	0.209
30	3	.472	.710	.844	-0.047	-0.014	0.023	0.097	0.159	0.459
30	4	.649	.817	.891	-0.024	-0.007	0.010	0.061	0.107	0.302
30	5	.789	.882	.923	-0.033	-0.004	0.005	0.032	0.066	0.142
40	3	.530	.753	.867	-0.046	-0.012	0.023	0.083	0.148	0.455
40	4	.698	.864	.924	-0.019	-0.006	0.010	0.048	0.095	0.293
40	5	.882	.930	.955	-0.017	-0.002	0.003	0.024	0.050	0.099
50	3	.570	.783	.876	-0.045	-0.011	0.025	0.068	0.140	0.454
50	4	.748	.896	.944	-0.016	-0.005	0.012	0.040	0.086	0.287
50	5	.915	.961	.975	-0.009	-0.001	0.001	0.017	0.036	0.069
60	3	.598	.804	.886	-0.044	-0.010	0.025	0.056	0.135	0.451
60	4	.787	.919	.957	-0.016	-0.004	0.012	0.033	0.080	0.284
60	5	.960	.983	.990	-0.004	0.000	0.001	0.011	0.023	0.042

Note. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criterion, they reflect the minimum and maximum value across replications. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table J6

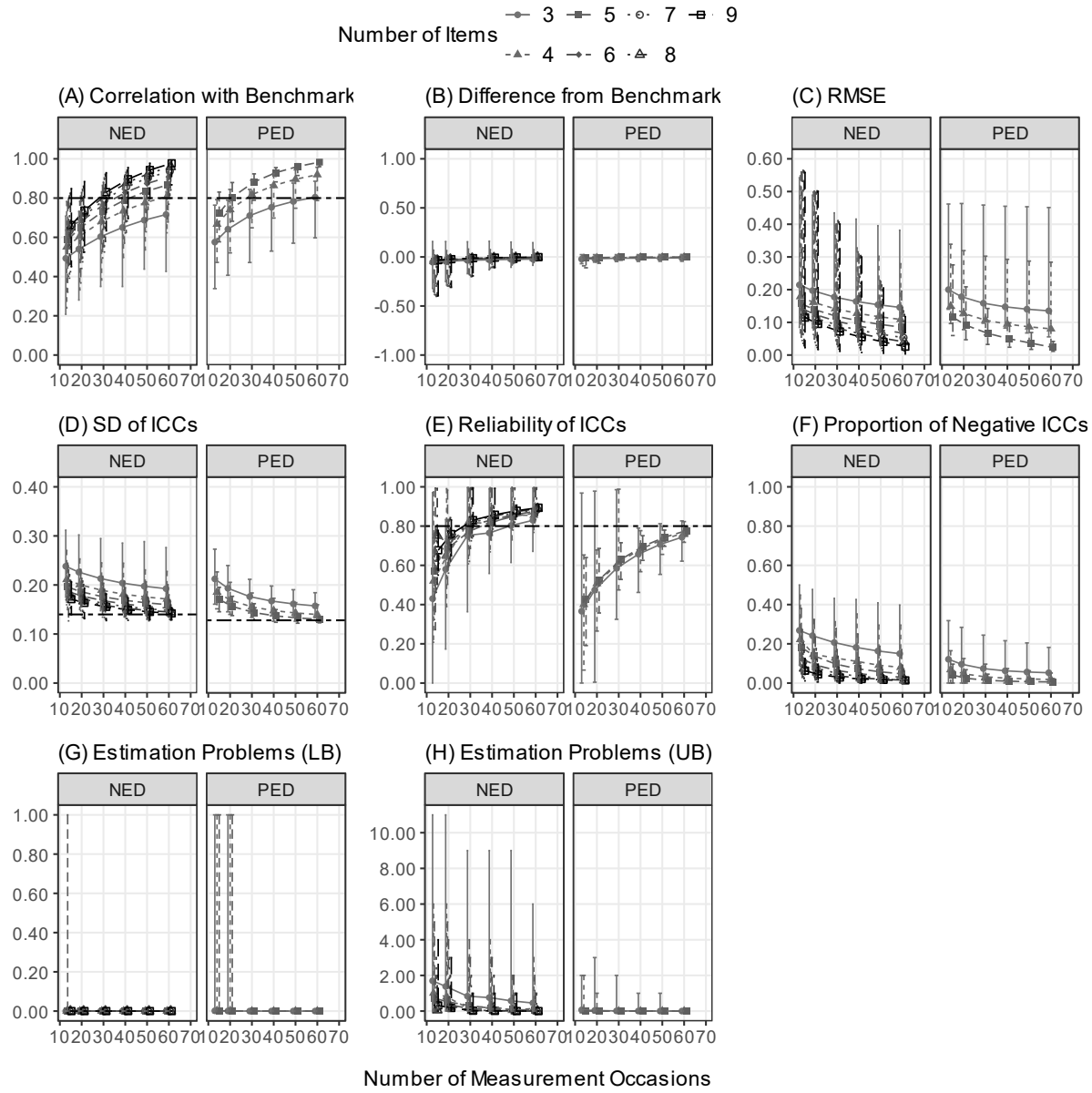
Absolute Evaluation Criteria for Raw Intraclass Correlations (ICCs) Aggregated Across Replications (Positive Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	0.161	0.212	0.273	.000	.366	.968	.000	.121	.318	0	0.002	1	0	0.066	2
14	4	0.151	0.185	0.227	.066	.401	.653	.000	.068	.165	0	0.001	1	0	0.014	2
14	5	0.145	0.171	0.195	.191	.425	.640	.000	.040	.097	0	0.002	1	0	0	0
20	3	0.145	0.193	0.240	.004	.474	.978	.000	.096	.284	0	0	1	0	0.034	3
20	4	0.140	0.169	0.205	.265	.507	.678	.000	.047	.125	0	0	1	0	0.003	1
20	5	0.138	0.156	0.177	.357	.526	.686	.000	.025	.068	0	0	1	0	0	0
30	3	0.143	0.176	0.212	.325	.585	.986	.000	.074	.244	0	0	0	0	0.010	2
30	4	0.134	0.155	0.184	.475	.614	.988	.000	.032	.097	0	0	0	0	0	0
30	5	0.128	0.144	0.162	.528	.630	.715	.000	.014	.045	0	0	0	0	0	0
40	3	0.137	0.167	0.198	.462	.657	.790	.000	.063	.216	0	0	0	0	0.004	1
40	4	0.126	0.148	0.173	.568	.683	.777	.000	.025	.074	0	0	0	0	0	0
40	5	0.124	0.137	0.150	.625	.697	.751	.000	.009	.034	0	0	0	0	0	0
50	3	0.135	0.161	0.190	.554	.706	.812	.000	.056	.205	0	0	0	0	0.001	1
50	4	0.126	0.143	0.164	.656	.730	.800	.000	.021	.062	0	0	0	0	0	0
50	5	0.122	0.133	0.142	.697	.743	.780	.000	.007	.023	0	0	0	0	0	0
60	3	0.133	0.157	0.184	.621	.743	.827	.000	.053	.182	0	0	0	0	0	0
60	4	0.126	0.140	0.160	.717	.766	.823	.000	.018	.051	0	0	0	0	0	0
60	5	0.123	0.130	0.135	.748	.777	.796	.000	.006	.023	0	0	0	0	0	0
70	5	-	0.128	-	-	.803	-	-	.006	-	-	0	-	-	0	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 5 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The minimum and maximum reflect the minimum and maximum value across replications. The reliability calculation required Fisher's *Z*-transformed ICCs; thus, participants affected by estimation problems were excluded from these analyses. The minimum sample size for the reliability calculation ranged from 173 to 176 across conditions. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. LB = lower bound estimation problems (ICCs ≤ -1/[k-1]). UB = upper bound estimation problems (ICCs ≥ 1). N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Figure J1

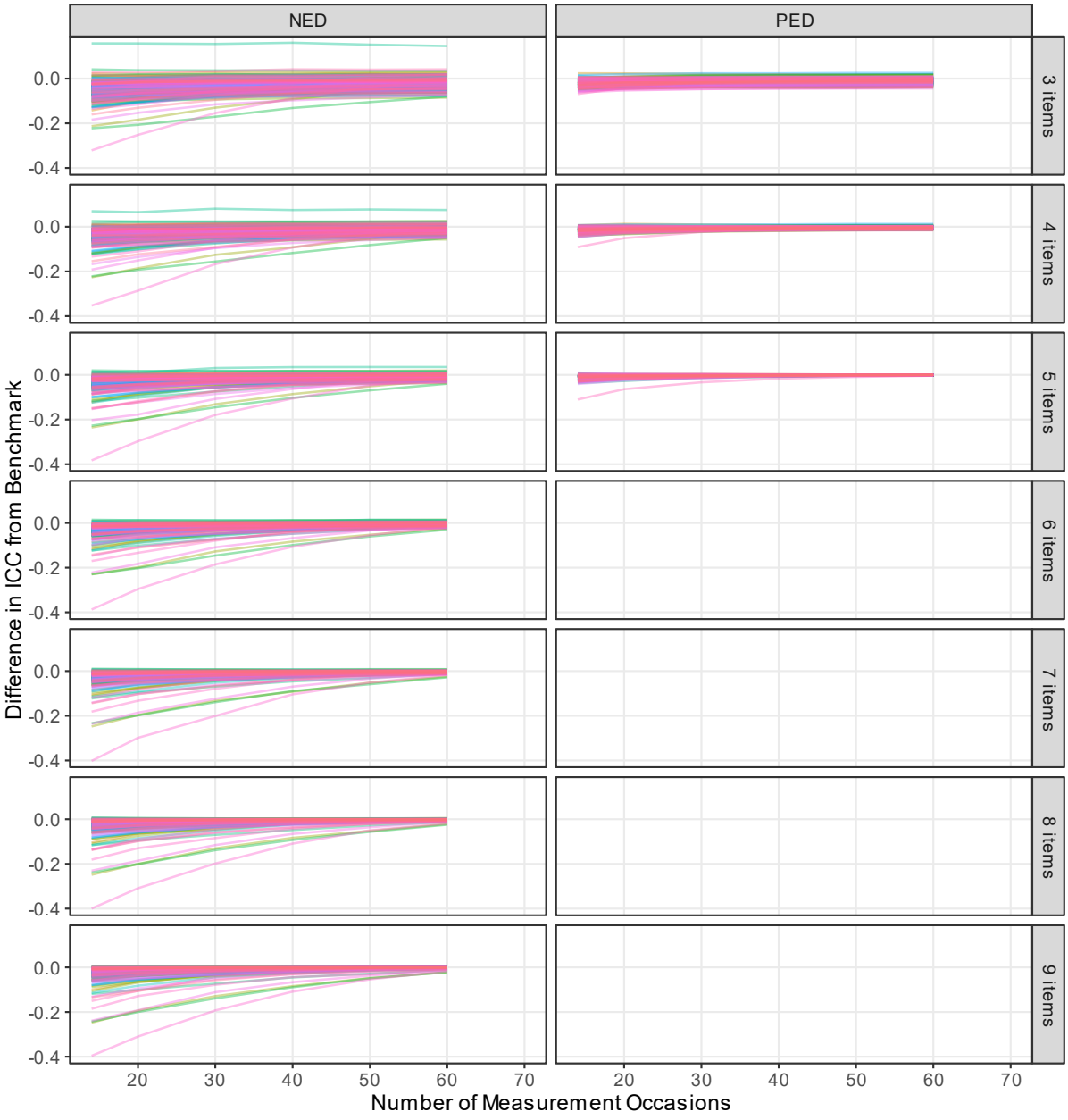
Results of the Evaluation Criteria for Raw Intraclass Correlations (ICCs) for Each Condition



Note. Panels (A) to (H) depict the evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (H) are absolute criteria. The panels present results for negative emotion differentiation (NED; left) and positive emotion differentiation (PED; right). The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across replications. The plots display mean values for each item number condition. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E), and the standard deviation (*SD*) of ICCs in the benchmark condition in (D). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs. LB = lower bound. UB = upper bound.

Figure J2

Person-Level Difference from Raw Benchmark Intraclass Correlations (ICC) Separate for Each Participant



Note. Each line represents a participant's person-level difference between the ICC in a given simulation condition and the ICC in the benchmark condition. The left column shows person-level differences negative emotion differentiation (NED), whereas the right column shows the person-level differences for positive emotion differentiation (PED). The rows display results for conditions using different numbers of items.

K: Results for Fisher's *Z*-Transformed Intraclass Correlations in the Replication Data Set

In the sensitivity analysis using Fisher's Z-transformed intraclass correlations, most evaluation criteria (except for number of estimation problems and raw intraclass correlations below zero) were based on Fisher's *Z*-transformed intraclass correlations. Thus, participants affected by estimation problems (i.e., intraclass correlations that fell outside the range of transformable values) were excluded from the corresponding analyses. Consequently, the root mean squared error (RMSE), which was calculated separately for each participant across replications within each condition, was based only on replications in which the respective participant was *not* excluded. Thus, the number of replications included in the RMSE calculation varied across participants and conditions. The minimum sample size and the number of replications used for the RMSE calculation are reported in the table notes.

Table K1

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Fisher's Z-Transformed Intraclass Correlations (ICCs) Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Negative Emotion Differentiation in the Replication Data Set)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
14	3	0.001	0.017	0.080	0.002	0.001	0.001	0.000	0.022
14	4	0.001	0.024	0.088	0.001	0.001	0.001	0.000	0.014
14	5	0.001	0.014	0.009	0.001	0.000	0.001	0.000	0.011
14	6	0.002	0.108	0.125	0.003	0.009	0.000	0.000	0.004
14	7	0.001	0.014	0.008	0.001	0.000	0.000	0.000	0.009
14	8	0.002	0.092	0.126	0.002	0.007	0.000	0.000	0.003
14	9	0.001	0.014	0.007	0.001	0.000	0.000	0.000	0.008
20	3	0.001	0.014	0.008	0.001	0.001	0.001	0.000	0.020
20	4	0.001	0.022	0.108	0.001	0.001	0.001	0.000	0.013
20	5	0.001	0.022	0.083	0.001	0.001	0.001	0.000	0.010
20	6	0.001	0.013	0.008	0.001	0.000	0.000	0.000	0.008
20	7	0.001	0.013	0.008	0.000	0.000	0.000	0.000	0.007
20	8	0.001	0.013	0.008	0.000	0.000	0.000	0.000	0.006
20	9	0.001	0.013	0.007	0.000	0.000	0.000	0.000	0.006
30	3	0.002	0.096	0.122	0.002	0.006	0.001	0.000	0.016
30	4	0.003	0.095	0.121	0.002	0.008	0.001	0.000	0.006
30	5	0.001	0.013	0.041	0.000	0.000	0.001	0.000	0.008
30	6	0.001	0.051	0.124	0.001	0.003	0.000	0.000	0.005
30	7	0.002	0.048	0.127	0.001	0.004	0.000	0.000	0.003
30	8	0.002	0.039	0.122	0.001	0.003	0.000	0.000	0.002
30	9	0.001	0.034	0.119	0.000	0.003	0.000	0.000	0.002

Table K1 (continued)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
40	3	0.001	0.027	0.104	0.001	0.002	0.001	0.000	0.016
40	4	0.002	0.060	0.123	0.001	0.005	0.001	0.000	0.008
40	5	0.001	0.033	0.124	0.000	0.002	0.000	0.000	0.006
40	6	0.001	0.010	0.039	0.000	0.000	0.000	0.000	0.004
40	7	0.001	0.009	0.009	0.000	0.000	0.000	0.000	0.003
40	8	0.001	0.009	0.009	0.000	0.000	0.000	0.000	0.002
40	9	0.001	0.012	0.106	0.000	0.001	0.000	0.000	0.001
50	3	0.001	0.011	0.064	0.001	0.001	0.001	0.000	0.014
50	4	0.003	0.087	0.120	0.001	0.007	0.001	0.000	0.003
50	5	0.001	0.028	0.122	0.000	0.002	0.000	0.000	0.004
50	6	0.000	0.008	0.008	0.000	0.000	0.000	0.000	0.003
50	7	0.001	0.015	0.117	0.000	0.001	0.000	0.000	0.001
50	8	0.000	0.006	0.009	0.000	0.000	0.000	0.000	0.001
50	9	0.000	0.006	0.009	0.000	0.000	0.000	0.000	0.000
60	3	0.001	0.018	0.096	0.001	0.001	0.001	0.000	0.012
60	4	0.001	0.029	0.119	0.001	0.003	0.001	0.000	0.006
60	5	0.001	0.018	0.117	0.000	0.002	0.000	0.000	0.003
60	6	0.000	0.011	0.113	0.000	0.001	0.000	0.000	0.002
60	7	0.000	0.005	0.091	0.000	0.000	0.000	0.000	0.000
60	8	0.000	0.004	0.008	0.000	0.000	0.000	0.000	0.000
60	9	0.000	0.003	0.008	0.000	0.000	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero. LB estimation probl.= lower bound estimation problems (ICCs ≤ -1/[k-1]). UB estimation probl. = upper bound estimation problems (ICCs ≥ 1).

Table K2

Relative Evaluation Criteria for Fisher's Z-Transformed Intraclass Correlations (ICCs) Aggregated Across Replications (Negative Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	.050	.458	.658	-1.028	-0.395	0.082	0.141	0.504	1.493
14	4	.247	.529	.734	-0.938	-0.333	0.070	0.125	0.444	1.860
14	5	.335	.573	.749	-0.890	-0.277	0.059	0.123	0.392	1.316
14	6	.106	.502	.767	-0.834	-0.175	3.208	0.118	0.476	8.118
14	7	.417	.623	.788	-0.823	-0.177	0.035	0.115	0.317	1.272
14	8	.121	.566	.815	-0.785	-0.098	2.128	0.115	0.384	6.727
14	9	.442	.655	.832	-0.741	-0.092	0.012	0.101	0.270	1.211
20	3	.260	.513	.709	-0.895	-0.381	0.082	0.139	0.475	1.241
20	4	.178	.590	.756	-0.774	-0.314	0.071	0.124	0.415	1.654
20	5	.361	.638	.802	-0.689	-0.254	0.059	0.122	0.358	1.636
20	6	.450	.674	.827	-0.652	-0.199	0.048	0.117	0.310	1.141
20	7	.470	.698	.830	-0.616	-0.150	0.034	0.107	0.272	1.116
20	8	.517	.717	.847	-0.605	-0.104	0.021	0.096	0.241	1.101
20	9	.526	.731	.881	-0.576	-0.062	0.010	0.072	0.219	1.074
30	3	.063	.548	.758	-0.629	-0.342	1.912	0.134	0.517	6.464
30	4	.019	.573	.813	-0.565	-0.256	2.553	0.125	0.492	7.178
30	5	.394	.723	.845	-0.499	-0.232	0.063	0.115	0.317	1.031
30	6	.321	.747	.873	-0.455	-0.170	0.434	0.104	0.291	3.513
30	7	.149	.766	.903	-0.445	-0.117	0.453	0.094	0.263	3.383
30	8	.165	.794	.913	-0.405	-0.073	0.294	0.078	0.220	2.782
30	9	.186	.818	.931	-0.356	-0.031	0.172	0.052	0.186	2.426

Table K2 (continued)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
40	3	.265	.634	.786	-0.619	-0.355	0.084	0.131	0.436	1.855
40	4	.193	.706	.850	-0.501	-0.275	0.611	0.123	0.394	4.168
40	5	.350	.779	.893	-0.420	-0.218	0.062	0.111	0.302	2.266
40	6	.433	.822	.910	-0.361	-0.162	0.049	0.099	0.237	0.785
40	7	.632	.852	.930	-0.295	-0.110	0.035	0.082	0.190	0.689
40	8	.688	.875	.938	-0.257	-0.063	0.020	0.063	0.152	0.665
40	9	.391	.892	.955	-0.208	-0.019	0.041	0.038	0.125	0.828
50	3	.073	.676	.811	-0.550	-0.350	0.087	0.132	0.417	0.838
50	4	.078	.714	.887	-0.446	-0.256	2.064	0.120	0.411	6.499
50	5	.349	.824	.910	-0.359	-0.209	0.064	0.108	0.288	1.947
50	6	.704	.869	.934	-0.284	-0.153	0.049	0.093	0.216	0.582
50	7	.375	.898	.948	-0.226	-0.101	0.035	0.073	0.172	1.076
50	8	.790	.923	.964	-0.172	-0.054	0.019	0.053	0.123	0.451
50	9	.797	.941	.976	-0.112	-0.011	0.013	0.027	0.089	0.441
60	3	.359	.705	.830	-0.514	-0.345	0.087	0.128	0.410	1.285
60	4	.333	.795	.906	-0.393	-0.269	0.078	0.118	0.346	2.063
60	5	.357	.860	.931	-0.306	-0.205	0.065	0.103	0.269	1.256
60	6	.406	.903	.951	-0.224	-0.147	0.051	0.088	0.202	0.755
60	7	.400	.934	.968	-0.160	-0.095	0.034	0.068	0.147	0.350
60	8	.846	.958	.981	-0.102	-0.048	0.017	0.044	0.099	0.291
60	9	.879	.976	.992	-0.047	-0.004	0.011	0.018	0.057	0.229

Note. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criterion, they reflect the minimum and maximum value across replications. The analyses (except for number of estimation problems and raw ICCs below zero) were based on Fisher's *Z*-transformed ICCs. Thus, participants affected by estimation problems were excluded from these analyses. The minimum sample size for the analyses ranged from 165 to 176 across conditions, and the number of replications used for the RMSE calculation ranged from 2,749 to 5,000 per participant. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table K3

Absolute Evaluation Criteria for Fisher's Z-Transformed Intraclass Correlations (ICCs) Aggregated Across Replications (Negative Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	0.239	0.343	1.461	.000	.430	.971	.091	.268	.500	0	0	0	0	1.698	11
14	4	0.266	0.351	1.478	.208	.520	.974	.085	.222	.449	0	0	0	0	1.010	6
14	5	0.278	0.353	0.468	.324	.572	.761	.057	.182	.381	0	0	1	0	0.721	5
14	6	0.263	0.837	2.836	.272	.757	.994	.028	.122	.273	0	0	0	0	0.086	2
14	7	0.302	0.370	0.471	.463	.638	.780	.034	.108	.205	0	0	0	0	0.451	3
14	8	0.310	0.703	2.440	.503	.752	.992	.023	.075	.165	0	0	0	0	0.054	2
14	9	0.324	0.384	0.461	.550	.677	.778	.011	.063	.125	0	0	0	0	0.315	4
20	3	0.225	0.323	0.502	.173	.580	.834	.068	.241	.477	0	0	0	0	1.382	11
20	4	0.247	0.334	2.010	.389	.646	.991	.040	.146	.324	0	0	0	0	0.798	6
20	5	0.265	0.340	1.467	.503	.686	.984	.040	.138	.324	0	0	0	0	0.522	5
20	6	0.276	0.344	0.445	.560	.713	.831	.017	.092	.233	0	0	0	0	0.367	4
20	7	0.285	0.351	0.444	.600	.733	.835	.017	.076	.176	0	0	0	0	0.299	3
20	8	0.290	0.357	0.451	.622	.748	.843	.006	.055	.119	0	0	0	0	0.230	2
20	9	0.299	0.363	0.432	.648	.760	.832	.006	.045	.097	0	0	0	0	0.175	3
30	3	0.206	0.514	2.781	.363	.753	.997	.040	.207	.432	0	0	0	0	0.835	9
30	4	0.247	0.687	2.781	.608	.819	.997	.011	.128	.312	0	0	0	0	0.159	3
30	5	0.243	0.318	1.444	.620	.773	.989	.017	.096	.278	0	0	0	0	0.312	4
30	6	0.262	0.381	2.029	.686	.802	.995	.006	.064	.170	0	0	0	0	0.128	3
30	7	0.276	0.410	2.040	.725	.818	.995	.006	.048	.136	0	0	0	0	0.046	2
30	8	0.281	0.393	1.994	.740	.825	.995	.000	.036	.080	0	0	0	0	0.029	1
30	9	0.283	0.377	1.994	.748	.830	.995	.000	.029	.062	0	0	0	0	0.021	1

Table K3 (continued)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
40	3	0.212	0.307	1.455	.558	.765	.991	.028	.182	.426	0	0	0	0	0.757	9
40	4	0.231	0.396	2.042	.669	.813	.996	.011	.110	.352	0	0	0	0	0.272	4
40	5	0.247	0.321	1.474	.728	.821	.992	.006	.073	.250	0	0	0	0	0.156	3
40	6	0.253	0.310	1.431	.752	.833	.992	.006	.048	.131	0	0	0	0	0.086	3
40	7	0.268	0.315	0.402	.785	.843	.905	.000	.034	.091	0	0	0	0	0.033	1
40	8	0.277	0.320	0.385	.803	.851	.898	.000	.026	.074	0	0	0	0	0.024	2
40	9	0.277	0.326	1.431	.806	.858	.993	.000	.021	.045	0	0	0	0	0.010	1
50	3	0.202	0.290	1.413	.614	.803	.992	.017	.164	.409	0	0	0	0	0.577	9
50	4	0.227	0.501	2.782	.729	.856	.998	.006	.091	.295	0	0	0	0	0.041	2
50	5	0.244	0.315	1.469	.780	.849	.994	.006	.058	.188	0	0	0	0	0.082	2
50	6	0.255	0.299	0.411	.807	.858	.926	.000	.037	.119	0	0	0	0	0.034	2
50	7	0.260	0.309	1.435	.820	.867	.994	.000	.026	.080	0	0	0	0	0.007	1
50	8	0.269	0.308	0.358	.834	.873	.907	.000	.019	.051	0	0	0	0	0.003	1
50	9	0.278	0.312	0.355	.848	.879	.907	.006	.016	.040	0	0	0	0	0.001	1
60	3	0.199	0.288	1.433	.671	.829	.994	.017	.150	.398	0	0	0	0	0.453	6
60	4	0.223	0.312	2.006	.768	.855	.997	.006	.080	.295	0	0	0	0	0.151	3
60	5	0.244	0.296	1.445	.818	.866	.995	.000	.047	.170	0	0	0	0	0.058	2
60	6	0.250	0.292	1.434	.834	.875	.995	.000	.030	.091	0	0	0	0	0.016	1
60	7	0.259	0.295	1.447	.849	.883	.995	.000	.020	.068	0	0	0	0	0.001	1
60	8	0.265	0.299	0.342	.858	.889	.915	.000	.015	.040	0	0	0	0	0	0
60	9	0.279	0.303	0.331	.875	.894	.911	.006	.014	.028	0	0	0	0	0	0
70	9	-	0.297	-	-	.906	-	-	.011	-	-	0	-	0	0	0

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 9 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The minimum and maximum reflect the minimum and maximum value across replications. The analyses (except for number of estimation problems and raw ICCs below zero) were based on Fisher's *Z*-transformed ICCs. Thus, participants affected by estimation problems were excluded from these analyses. The minimum sample size for the analyses ranged from 165 to 176 across conditions. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with a raw ICC below zero. LB = lower bound estimation problems (ICCs ≤ -1/[k-1]). UB = upper bound estimation problems (ICCs ≥ 1). N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table K4

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Fisher's Z-Transformed Intraclass Correlations (ICCs) Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Positive Emotion Differentiation in the Replication Data Set)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
14	3	0.001	0.018	0.112	0.002	0.001	0.001	0.001	0.004
14	4	0.001	0.01	0.008	0.001	0.000	0.000	0.000	0.002
14	5	0.001	0.007	0.009	0.001	0.000	0.000	0.001	0.000
20	3	0.001	0.012	0.07	0.001	0.001	0.001	0.000	0.003
20	4	0.001	0.009	0.007	0.001	0.000	0.000	0.000	0.001
20	5	0.000	0.005	0.007	0.001	0.000	0.000	0.000	0.000
30	3	0.001	0.012	0.085	0.001	0.001	0.001	0.000	0.001
30	4	0.001	0.008	0.092	0.001	0.000	0.000	0.000	0.000
30	5	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.000
40	3	0.001	0.011	0.004	0.001	0.000	0.001	0.000	0.001
40	4	0.001	0.008	0.005	0.001	0.000	0.000	0.000	0.000
40	5	0.000	0.002	0.003	0.000	0.000	0.000	0.000	0.000
50	3	0.001	0.011	0.004	0.001	0.000	0.001	0.000	0.000
50	4	0.001	0.008	0.005	0.000	0.000	0.000	0.000	0.000
50	5	0.000	0.002	0.002	0.000	0.000	0.000	0.000	0.000
60	3	0.001	0.011	0.003	0.001	0.000	0.001	0.000	0.000
60	4	0.001	0.008	0.005	0.000	0.000	0.000	0.000	0.000
60	5	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero. LB estimation probl.= lower bound estimation problems (ICCs ≤ -1/[k-1]). UB estimation probl. = upper bound estimation problems (ICCs ≥ 1).

Table K5

Relative Evaluation Criteria for Fisher's Z-Transformed Intraclass Correlations (ICCs) Aggregated Across Replications (Positive Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	.100	.568	.767	-0.254	-0.176	-0.021	0.213	0.354	1.275
14	4	.463	.662	.787	-0.214	-0.097	-0.021	0.183	0.265	0.692
14	5	.567	.720	.828	-0.171	-0.026	0.014	0.097	0.207	0.492
20	3	.265	.635	.796	-0.242	-0.170	-0.013	0.181	0.315	0.862
20	4	.511	.737	.840	-0.143	-0.088	-0.016	0.147	0.226	0.631
20	5	.649	.802	.882	-0.095	-0.016	0.011	0.076	0.160	0.366
30	3	.268	.703	.851	-0.236	-0.165	-0.010	0.146	0.288	0.821
30	4	.291	.814	.892	-0.114	-0.082	-0.015	0.115	0.192	0.577
30	5	.788	.881	.923	-0.047	-0.008	0.009	0.052	0.115	0.247
40	3	.503	.746	.865	-0.232	-0.163	-0.013	0.127	0.272	0.799
40	4	.671	.860	.922	-0.111	-0.079	-0.013	0.091	0.172	0.554
40	5	.881	.929	.955	-0.023	-0.004	0.005	0.038	0.085	0.172
50	3	.545	.776	.880	-0.230	-0.162	-0.010	0.114	0.263	0.789
50	4	.724	.893	.943	-0.104	-0.077	-0.014	0.077	0.159	0.541
50	5	.913	.961	.975	-0.012	-0.002	0.001	0.028	0.062	0.121
60	3	.587	.797	.890	-0.230	-0.161	-0.010	0.105	0.256	0.778
60	4	.758	.916	.956	-0.103	-0.076	-0.012	0.065	0.150	0.535
60	5	.961	.983	.990	-0.005	-0.001	0.001	0.018	0.040	0.074

Note. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criterion, they reflect the minimum and maximum value across replications. The analyses were based on Fisher's *Z*-transformed ICCs. Thus, participants affected by estimation problems were excluded from these analyses. The minimum sample size for the analyses ranged from 173 to 176 across conditions, and the number of replications used for the RMSE calculation ranged from 4,775 to 5,000 per participant. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table K6

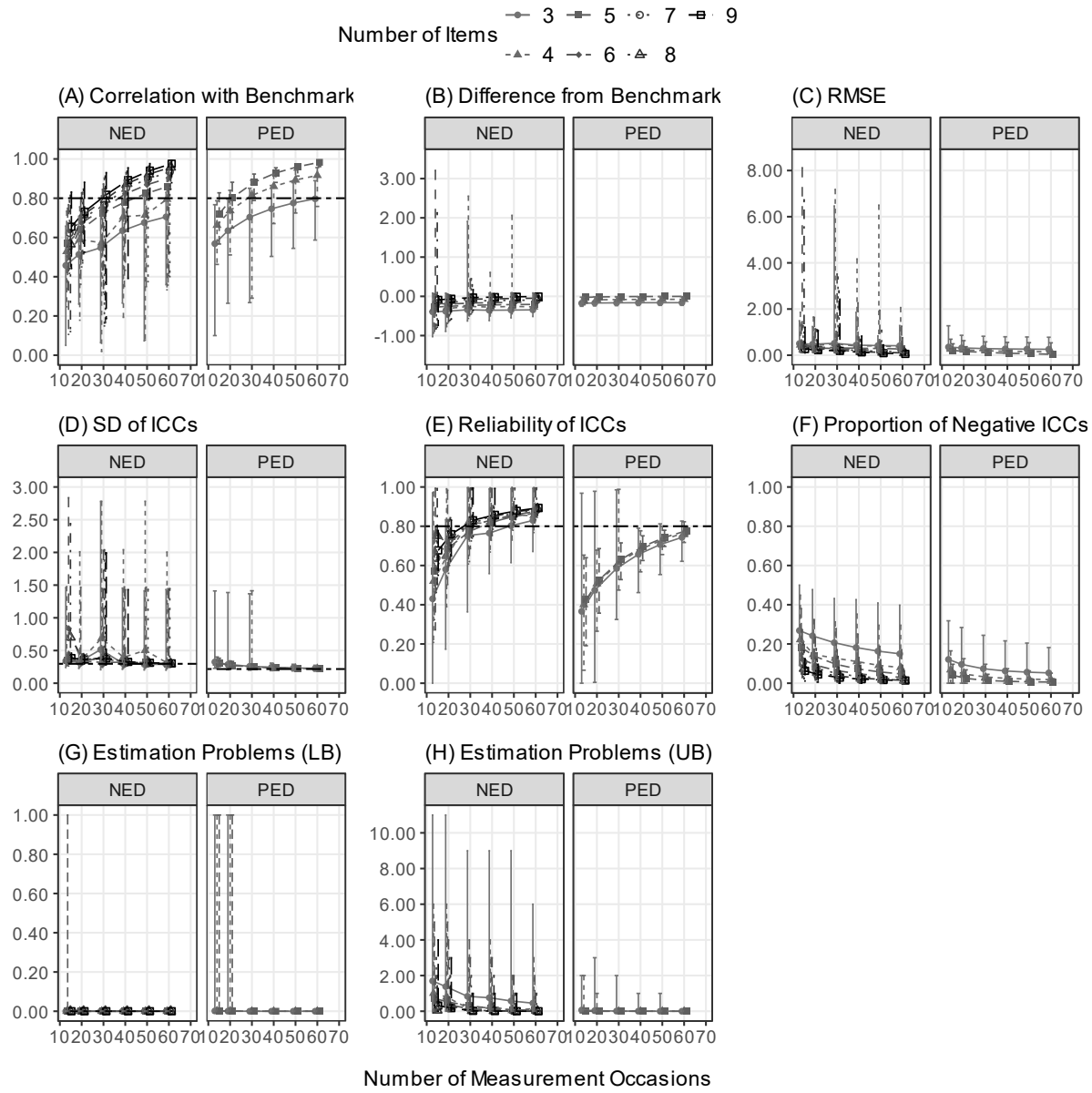
Absolute Evaluation Criteria for Fisher's Z-Transformed Intraclass Correlations (ICCs) Aggregated Across Replications (Positive Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	0.227	0.325	1.411	.000	.366	.968	.000	.121	.318	0	0.002	1	0	0.066	2
14	4	0.245	0.308	0.401	.066	.401	.653	.000	.068	.165	0	0.001	1	0	0.014	2
14	5	0.254	0.303	0.381	.191	.425	.640	.000	.040	.097	0	0.002	1	0	0	0
20	3	0.205	0.287	1.387	.004	.474	.978	.000	.096	.284	0	0	1	0	0.034	3
20	4	0.225	0.277	0.340	.265	.507	.678	.000	.047	.125	0	0	1	0	0.003	1
20	5	0.233	0.272	0.334	.357	.526	.686	.000	.025	.068	0	0	1	0	0	0
30	3	0.200	0.259	1.368	.325	.585	.986	.000	.074	.244	0	0	0	0	0.01	2
30	4	0.213	0.251	1.413	.475	.614	.988	.000	.032	.097	0	0	0	0	0	0
30	5	0.218	0.247	0.281	.528	.630	.715	.000	.014	.045	0	0	0	0	0	0
40	3	0.192	0.244	0.307	.462	.657	.790	.000	.063	.216	0	0	0	0	0.004	1
40	4	0.202	0.237	0.281	.568	.683	.777	.000	.025	.074	0	0	0	0	0	0
40	5	0.210	0.234	0.258	.625	.697	.751	.000	.009	.034	0	0	0	0	0	0
50	3	0.188	0.234	0.289	.554	.706	.812	.000	.056	.205	0	0	0	0	0.001	1
50	4	0.201	0.229	0.265	.656	.730	.800	.000	.021	.062	0	0	0	0	0	0
50	5	0.208	0.226	0.244	.697	.743	.780	.000	.007	.023	0	0	0	0	0	0
60	3	0.185	0.228	0.274	.621	.743	.827	.000	.053	.182	0	0	0	0	0	0
60	4	0.202	0.223	0.255	.717	.766	.823	.000	.018	.051	0	0	0	0	0	0
60	5	0.207	0.220	0.231	.748	.777	.796	.000	.006	.023	0	0	0	0	0	0
70	5	-	0.217	-	-	.803	-	-	.006	-	-	0	-	-	0	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 5 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The minimum and maximum reflect the minimum and maximum value across replications. The analyses (except for number of estimation problems and raw ICCs below zero) were based on Fisher's *Z*-transformed ICCs. Thus, participants affected by estimation problems were excluded from these analyses. The minimum sample size for the analyses ranged from 173 to 176 across conditions. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with a raw ICC below zero. LB = lower bound estimation problems (ICCs ≤ -1/[k-1]). UB = upper bound estimation problems (ICCs ≥ 1). N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Figure K1

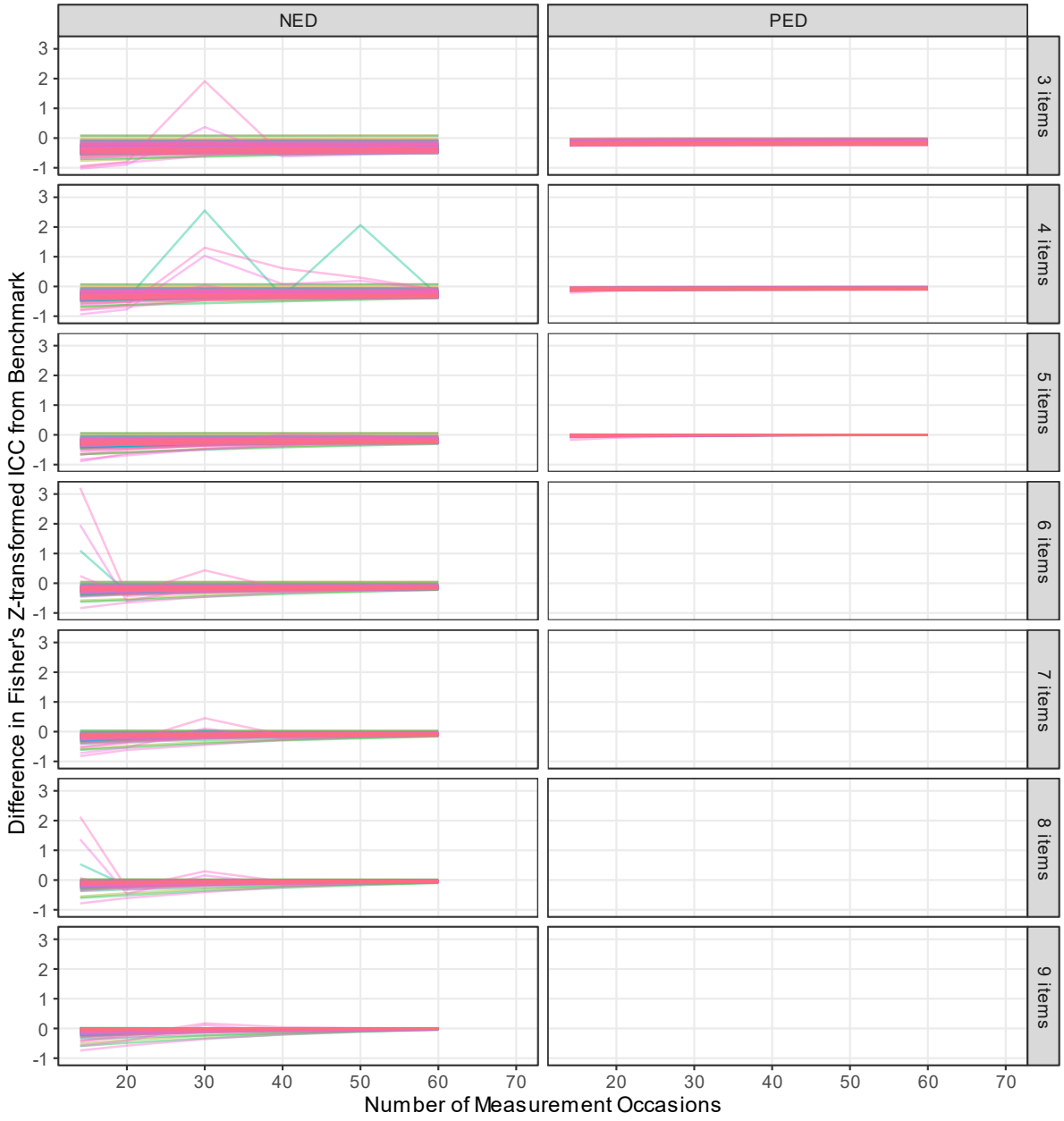
Results of the Evaluation Criteria for Fisher's Z-Transformed Intraclass Correlations (ICCs) for Each Condition



Note. Panels (A) to (H) depict the evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (H) are absolute criteria. The panels present results for negative emotion differentiation (NED; left) and positive emotion differentiation (PED; right). The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across replications. The plots display mean values for each item number condition. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E), and the standard deviation (*SD*) of ICCs in the benchmark condition in (D). The proportion of negative ICCs and estimation problems use raw ICCs. LB = lower bound. UB = upper bound.

Figure K2

Person-Level Difference from Fisher's Z-Transformed Benchmark Intraclass Correlations (ICC) Separate for Each Participant



Note. Each line represents a participant's person-level difference between the ICC in a given simulation condition and the ICC in the benchmark condition. The left column shows person-level differences negative emotion differentiation (NED), whereas the right column shows the person-level differences for positive emotion differentiation (PED). The rows display results for conditions using different numbers of items.

L: Sensitivity Analyses Drawing Measurement Occasions in Temporal Order in the Replication Data Set

Since there is only one possible temporal ordering of measurement occasions, the re-drawing of measurement occasions in case of zero variance (as implemented in the main analysis; see Footnote 2 in the manuscript) would not make sense. Thus, for the sensitivity analysis using measurement occasions drawn by temporal order, participants with zero variance across emotion items and measurement occasions were excluded from the corresponding analyses because intraclass correlation coefficients could not be calculated. Consequently, the sample size varied across conditions. The number of excluded participants and the minimum sample sizes across conditions are reported in the corresponding table notes. Furthermore, the proportion of negative intraclass correlations was calculated as the number of participants with negative intraclass correlations divided by the number of participants for whom an intraclass correlation could be calculated (i.e., participants with zero variance were excluded from the denominator).

Table L1

Relative Evaluation Criteria from the Sensitivity Analysis for Raw Intraclass Correlations (ICCs) Using Measurement Occasions Drawn by Temporal Order (Negative Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max
14	3	.448	.562	.656	-0.383	-0.049	0.545
14	4	.539	.626	.698	-0.384	-0.041	0.513
14	5	.600	.667	.715	-0.383	-0.035	0.489
14	6	.649	.695	.732	-0.382	-0.031	0.469
14	7	.684	.716	.744	-0.380	-0.028	0.453
14	8	.711	.731	.744	-0.378	-0.026	0.44
14	9	-	.742	-	-0.377	-0.024	0.429
20	3	.446	.629	.728	-0.379	-0.036	0.471
20	4	.568	.696	.763	-0.378	-0.027	0.438
20	5	.643	.739	.783	-0.376	-0.022	0.414
20	6	.713	.768	.799	-0.373	-0.018	0.395
20	7	.749	.790	.817	-0.371	-0.016	0.380
20	8	.787	.805	.821	-0.369	-0.014	0.367
20	9	-	.818	-	-0.366	-0.012	0.357
30	3	.509	.667	.760	-0.320	-0.032	0.474
30	4	.629	.741	.812	-0.313	-0.024	0.441
30	5	.692	.789	.842	-0.309	-0.019	0.416
30	6	.763	.823	.862	-0.305	-0.016	0.397
30	7	.807	.847	.879	-0.303	-0.014	0.381
30	8	.849	.866	.881	-0.302	-0.012	0.369
30	9	-	.881	-	-0.301	-0.011	0.358
40	3	.531	.691	.804	-0.335	-0.027	0.462
40	4	.655	.770	.855	-0.331	-0.019	0.424
40	5	.722	.821	.872	-0.328	-0.014	0.394
40	6	.792	.857	.897	-0.326	-0.011	0.371
40	7	.831	.884	.913	-0.325	-0.008	0.352
40	8	.879	.904	.914	-0.324	-0.007	0.337
40	9	-	.921	-	-0.324	-0.005	0.324

Table L1 (continued)

Condition		Correlation			Difference		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max
50	3	.518	.720	.805	-0.334	-0.025	0.181
50	4	.670	.804	.874	-0.329	-0.018	0.159
50	5	.762	.857	.901	-0.326	-0.013	0.163
50	6	.811	.894	.925	-0.323	-0.01	0.166
50	7	.881	.920	.945	-0.322	-0.007	0.167
50	8	.928	.939	.948	-0.321	-0.006	0.168
50	9	-	.954	-	-0.321	-0.004	0.169
60	3	.509	.728	.835	-0.236	-0.022	0.182
60	4	.664	.819	.903	-0.224	-0.014	0.119
60	5	.765	.877	.927	-0.216	-0.01	0.123
60	6	.818	.917	.946	-0.210	-0.007	0.126
60	7	.901	.945	.966	-0.205	-0.004	0.128
60	8	.953	.966	.976	-0.202	-0.003	0.130
60	9	-	.982	-	-0.199	-0.001	0.131
70	3	.511	.737	.827	-0.085	-0.02	0.143
70	4	.678	.832	.919	-0.057	-0.013	0.073
70	5	.778	.892	.939	-0.037	-0.008	0.033
70	6	.834	.933	.959	-0.023	-0.005	0.015
70	7	.925	.962	.980	-0.013	-0.003	0.010
70	8	.974	.983	.988	-0.006	-0.001	0.005

Note. Since there is only one possible temporal ordering of measurement occasions, participants with zero variance across emotion items and measurement occasions were excluded from corresponding analyses. This occurred more frequently in conditions with fewer items and fewer measurement occasions. The maximum number of participants excluded due to variance-related problems ranged from 0 to 46 across conditions. The minimum sample size for analyses ranged from 130 to 176 across conditions. For the ordered draws of measurement occasions, we examined all possible item sets for each condition (with one replication each); thus, the minimum and maximum values reflect minimum and maximum values across the possible item sets. Since there was only one item set containing the maximum number of items, minimum and maximum values do not apply here. For the difference in ICCs, these values reflect the minimum and maximum value of the person-level average difference in ICCs across item sets. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table L2

Absolute Evaluation Criteria from the Sensitivity Analysis for Raw Intraclass Correlations (ICCs) Using Measurement Occasions Drawn by Temporal Order (Negative Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	0.221	0.251	0.326	.217	.501	.700	.111	.246	.382	0	0	0	1	2.952	9
14	4	0.205	0.226	0.259	.421	.601	.977	.097	.196	.391	0	0	0	1	2.286	6
14	5	0.192	0.213	0.233	.491	.634	.733	.080	.162	.302	0	0	0	1	1.952	4
14	6	0.186	0.204	0.216	.976	.984	.992	.045	.118	.229	0	0	0	0	0.012	1
14	7	0.186	0.198	0.205	.653	.691	.734	.080	.114	.165	0	0	0	1	1.444	3
14	8	0.187	0.193	0.199	.977	.980	.988	.062	.080	.108	0	0	0	0	0	0
14	9	-	0.190	-	-	.726	-	-	.068	-	-	0	-	-	1	-
20	3	0.209	0.240	0.305	.474	.649	.810	.114	.228	.359	0	0	0	0	1.762	7
20	4	0.198	0.218	0.257	.630	.756	.984	.080	.148	.253	0	0	0	0	1	3
20	5	0.187	0.206	0.233	.692	.760	.983	.062	.131	.259	0	0	0	0	0.706	3
20	6	0.180	0.198	0.214	.749	.786	.984	.045	.093	.176	0	0	0	0	0.429	2
20	7	0.180	0.193	0.200	.771	.800	.826	.045	.084	.136	0	0	0	0	0.25	1
20	8	0.182	0.189	0.193	.800	.811	.827	.051	.064	.091	0	0	0	0	0.111	1
20	9	-	0.186	-	-	.821	-	-	.057	-	-	0	-	-	0	-
30	3	0.183	0.218	0.288	.617	.775	.988	.074	.196	.378	0	0	0	0	0.702	5
30	4	0.176	0.197	0.230	.694	.830	.989	.040	.119	.273	0	0	0	0	0.127	3
30	5	0.169	0.186	0.206	.750	.803	.989	.034	.091	.210	0	0	0	0	0.183	3
30	6	0.164	0.178	0.194	.776	.818	.863	.028	.063	.131	0	0	0	0	0.083	1
30	7	0.164	0.173	0.185	.807	.833	.990	.028	.050	.102	0	0	0	0	0	0
30	8	0.164	0.170	0.174	.826	.838	.853	.028	.038	.057	0	0	0	0	0	0
30	9	-	0.167	-	-	.844	-	-	.028	-	-	0	-	-	0	-
40	3	0.173	0.208	0.280	.678	.784	.991	.051	.174	.378	0	0	0	0	0.702	6
40	4	0.165	0.187	0.224	.756	.823	.992	.034	.105	.286	0	0	0	0	0.294	2
40	5	0.162	0.175	0.199	.795	.834	.903	.023	.067	.199	0	0	0	0	0.175	2
40	6	0.158	0.168	0.187	.820	.846	.890	.017	.044	.108	0	0	0	0	0.083	1
40	7	0.157	0.163	0.175	.836	.854	.882	.017	.031	.074	0	0	0	0	0.028	1
40	8	0.156	0.159	0.164	.852	.861	.874	.017	.025	.028	0	0	0	0	0	0
40	9	-	0.156	-	-	.865	-	-	.023	-	-	0	-	-	0	-

Table L2 (continued)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
50	3	0.168	0.200	0.270	.728	.814	.905	.023	.165	.369	0	0	0	0	0.452	4
50	4	0.158	0.180	0.219	.785	.859	.996	.023	.090	.284	0	0	0	0	0.016	1
50	5	0.157	0.169	0.193	.827	.857	.921	.017	.055	.176	0	0	0	0	0.048	1
50	6	0.154	0.162	0.179	.847	.867	.913	.011	.034	.080	0	0	0	0	0.012	1
50	7	0.153	0.157	0.167	.861	.875	.903	.011	.023	.051	0	0	0	0	0	0
50	8	0.151	0.154	0.159	.875	.881	.893	.006	.014	.023	0	0	0	0	0	0
50	9	-	0.152	-	-	.886	-	-	.006	-	-	0	-	-	0	-
60	3	0.157	0.193	0.261	.742	.833	.917	.023	.148	.352	0	0	0	0	0.381	4
60	4	0.149	0.172	0.213	.797	.856	.994	.011	.077	.261	0	0	0	0	0.119	1
60	5	0.147	0.161	0.189	.833	.867	.927	.006	.043	.136	0	0	0	0	0.048	1
60	6	0.145	0.154	0.175	.853	.877	.923	.006	.026	.068	0	0	0	0	0.012	1
60	7	0.143	0.150	0.162	.869	.884	.915	.006	.017	.051	0	0	0	0	0	0
60	8	0.143	0.146	0.153	.883	.890	.899	.000	.011	.017	0	0	0	0	0	0
60	9	-	0.144	-	-	.895	-	-	.011	-	-	0	-	-	0	-
70	3	0.154	0.189	0.261	.767	.853	.995	.017	.138	.358	0	0	0	0	0.321	4
70	4	0.145	0.168	0.216	.818	.880	.995	.011	.071	.261	0	0	0	0	0.016	1
70	5	0.143	0.157	0.193	.852	.885	.996	.006	.039	.131	0	0	0	0	0	0
70	6	0.140	0.150	0.170	.869	.889	.931	.006	.025	.074	0	0	0	0	0.012	1
70	7	0.139	0.145	0.158	.882	.896	.923	.000	.016	.045	0	0	0	0	0	0
70	8	0.139	0.142	0.148	.896	.901	.909	.000	.012	.017	0	0	0	0	0	0
70	9	-	0.140	-	-	.906	-	-	.011	-	-	0	-	-	0	-

Note. Since there is only one possible temporal ordering of measurement occasions, participants with zero variance across emotion items and measurement occasions were excluded from corresponding analyses. This occurred more frequently in conditions with fewer items and fewer measurement occasions. The maximum number of participants excluded due to variance-related problems ranged from 0 to 46 across conditions. The minimum sample size for analyses ranged from 130 to 176 across conditions. Since reliability calculation required Fisher's *Z*-transformed ICCs, participants affected by estimation problems were additionally excluded from these analyses. The minimum sample size ranged from 121 to 176 across conditions for the reliability analysis. For the ordered draws of measurement occasions, we examined all possible item sets for each condition (with one replication each); thus, the minimum and maximum values reflect minimum and maximum values across the possible item sets. Since there was only one item set containing the maximum number of items, minimum and maximum values do not apply here. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. LB = lower bound estimation problems. UB = upper bound estimation problems. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table L3

Relative Evaluation Criteria from the Sensitivity Analysis for Raw Intraclass Correlations (ICCs) Using Measurement Occasions Drawn by Temporal Order (Positive Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max
14	3	.361	.483	.565	-0.346	-0.017	0.419
14	4	.520	.568	.610	-0.339	-0.010	0.408
14	5	-	.625	.-	-0.335	-0.005	0.402
20	3	.423	.604	.696	-0.307	-0.020	0.247
20	4	.585	.701	.754	-0.303	-0.014	0.255
20	5	-	.759	-	-0.301	-0.010	0.259
30	3	.507	.683	.771	-0.226	-0.014	0.183
30	4	.673	.789	.830	-0.234	-0.007	0.182
30	5	-	.851	-	-0.237	-0.004	0.182
40	3	.546	.738	.798	-0.152	-0.010	0.112
40	4	.743	.851	.878	-0.158	-0.004	0.112
40	5	-	.919	-	-0.162	0.000	0.113
50	3	.571	.768	.822	-0.170	-0.010	0.078
50	4	.774	.884	.913	-0.173	-0.003	0.088
50	5	-	.951	-	-0.175	0.000	0.094
60	3	.610	.805	.865	-0.108	-0.006	0.075
60	4	.809	.921	.947	-0.095	0.000	0.085
60	5	-	.983	-	-0.090	0.003	0.090
70	3	.642	.821	.879	-0.044	-0.010	0.025
70	4	.826	.937	.964	-0.016	-0.003	0.013

Note. Since there is only one possible temporal ordering of measurement occasions, participants with zero variance across emotion items and measurement occasions were excluded from corresponding analyses. The maximum number of participants excluded due to variance-related problems ranged from 0 to 1 across conditions. The minimum sample size for analyses ranged from 175 to 176 across conditions. For the ordered draws of measurement occasions, we examined all possible item sets for each condition (with one replication each); thus, the minimum and maximum values reflect minimum and maximum values across the possible item sets. Since there was only one item set containing the maximum number of items, minimum and maximum values do not apply here. For the difference in ICCs, the values reflect the minimum and maximum value of the person-level average difference in ICCs across item sets. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table L4

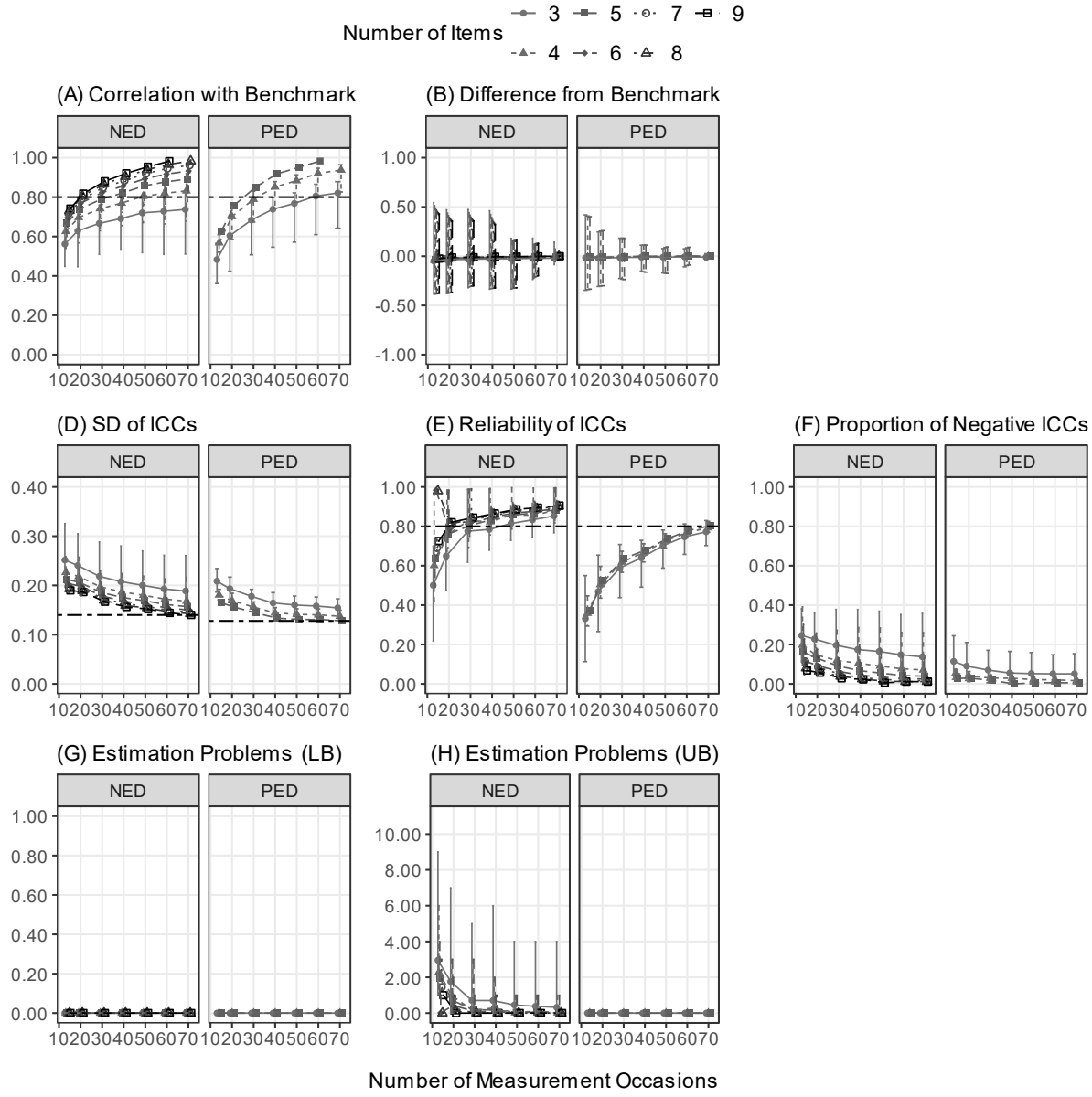
Absolute Evaluation Criteria from the Sensitivity Analysis for Raw Intraclass Correlations (ICCs) Using Measurement Occasions Drawn by Temporal Order (Positive Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	0.186	0.209	0.234	.113	.331	.550	.023	.114	.244	0	0	0	0	0	0
14	4	0.170	0.180	0.192	.293	.358	.447	.028	.056	.074	0	0	0	0	0	0
14	5	-	0.165	-	-	.371	-	-	.028	-	-	0	-	-	0	-
20	3	0.168	0.193	0.217	.266	.469	.654	.017	.090	.210	0	0	0	0	0	0
20	4	0.158	0.169	0.184	.436	.504	.597	.011	.040	.062	0	0	0	0	0	0
20	5	-	0.156	-	-	.526	-	-	.028	-	-	0	-	-	0	-
30	3	0.155	0.178	0.192	.438	.588	.708	.006	.070	.170	0	0	0	0	0	0
30	4	0.146	0.156	0.166	.567	.618	.673	.006	.030	.045	0	0	0	0	0	0
30	5	-	0.144	-	-	.637	-	-	.017	-	-	0	-	-	0	-
40	3	0.141	0.164	0.185	.491	.640	.729	.000	.055	.165	0	0	0	0	0	0
40	4	0.135	0.144	0.155	.619	.666	.711	.000	.028	.057	0	0	0	0	0	0
40	5	-	0.133	-	-	.680	-	-	.000	-	-	0	-	-	0	-
50	3	0.139	0.160	0.178	.588	.702	.781	.000	.052	.159	0	0	0	0	0	0
50	4	0.133	0.141	0.153	.690	.725	.762	.000	.022	.034	0	0	0	0	0	0
50	5	-	0.131	-	-	.737	-	-	.006	-	-	0	-	-	0	-
60	3	0.139	0.158	0.176	.657	.747	.812	.000	.051	.148	0	0	0	0	0	0
60	4	0.132	0.140	0.152	.742	.768	.801	.000	.020	.040	0	0	0	0	0	0
60	5	-	0.131	-	-	.779	-	-	.006	-	-	0	-	-	0	-
70	3	0.138	0.154	0.173	.702	.772	.829	.000	.051	.153	0	0	0	0	0	0
70	4	0.130	0.137	0.150	.770	.793	.825	.000	.016	.023	0	0	0	0	0	0
70	5	-	0.128	-	-	.803	-	-	.006	-	-	0	-	-	0	-

Note. Since there is only one possible temporal ordering of measurement occasions, participants with zero variance across emotion items and measurement occasions were excluded from corresponding analyses. The maximum number of participants excluded due to variance-related problems ranged from 0 to 1 across conditions. The minimum sample size for analyses ranged from 175 to 176 across conditions. For the ordered draws of measurement occasions, we examined all possible item sets for each condition (with one replication each); thus, the minimum and maximum values reflect minimum and maximum values across the possible item sets. Since there was only one item set containing the maximum number of items, minimum and maximum values do not apply here. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. LB = lower bound estimation problems. UB = upper bound estimation problems. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Figure L1

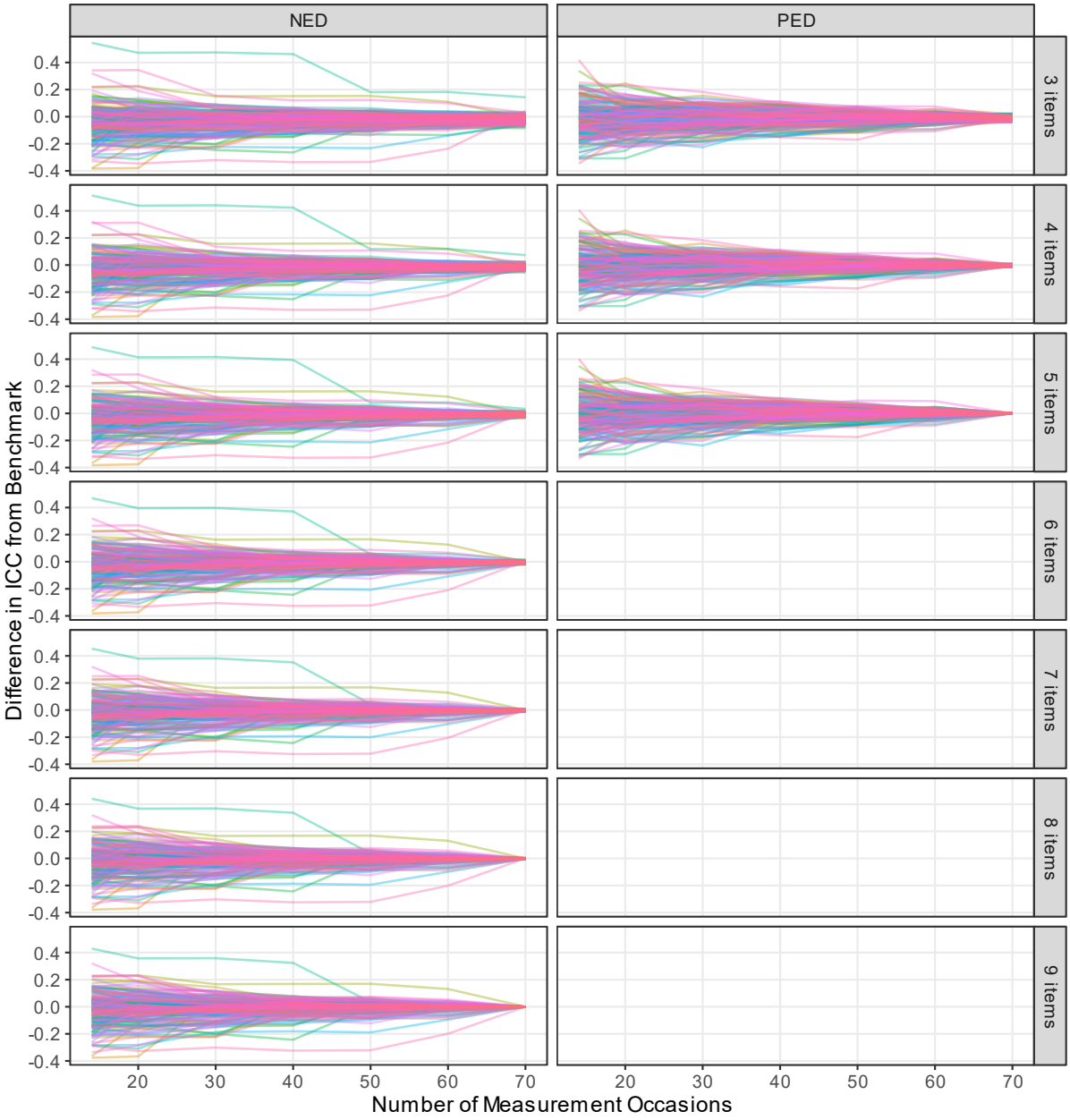
Results of the Evaluation Criteria for Raw Intraclass Correlations (ICCs) for Each Condition



Note. Panels (A) to (F) depict the evaluation criteria: (A) to (B) are relative to the benchmark condition; (D) to (F) are absolute criteria. The panels present results for negative emotion differentiation (NED; left) and positive emotion differentiation (PED; right). For the ordered draws of measurement occasions, we examined all possible item sets for each condition (with one replication each); thus, the minimum and maximum values reflect minimum and maximum values across the possible item sets. For the difference in ICCs, these values reflect the minimum and maximum value of the person-level average difference in ICCs across item sets. Since there was only one item set containing the maximum number of items, minimum and maximum values do not apply here. The RMSE was calculated per participant across replications and, thus, does not apply to the ordered draws of measurement occasions. The reliability calculation required Fisher's *Z*-transformed ICCs. LB = lower bound. UB = upper bound.

Figure L2

Person-Level Difference from Raw Benchmark Intraclass Correlations (ICC) Separate for Each Participant



Note. Each line represents a participant's person-level difference between the ICC in a given simulation condition and the ICC in the benchmark condition. The left column shows person-level differences negative emotion differentiation (NED), whereas the right column shows the person-level differences for positive emotion differentiation (PED). The rows display results for conditions using different numbers of items.

M: Sensitivity Analyses Using Different Treatments of Negative Intraclass Correlations in the Replication Data Set

Table M1

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Setting Negative ICCs to Zero Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Negative Emotion Differentiation in the Replication Data Set)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
14	3	0.001	0.006	0.004	0.002	0.000	0.001	0.000	0.021
14	4	0.001	0.006	0.004	0.002	0.000	0.001	0.000	0.014
14	5	0.001	0.006	0.003	0.001	0.000	0.001	0.000	0.011
14	6	0.001	0.006	0.003	0.003	0.000	0.000	0.000	0.004
14	7	0.001	0.006	0.003	0.001	0.000	0.000	0.000	0.009
14	8	0.001	0.006	0.003	0.002	0.000	0.000	0.000	0.003
14	9	0.001	0.006	0.003	0.001	0.000	0.000	0.000	0.008
20	3	0.001	0.006	0.004	0.002	0.000	0.001	0.000	0.020
20	4	0.001	0.006	0.004	0.001	0.000	0.001	0.000	0.013
20	5	0.001	0.006	0.004	0.001	0.000	0.001	0.000	0.010
20	6	0.001	0.006	0.003	0.001	0.000	0.000	0.000	0.008
20	7	0.001	0.006	0.003	0.000	0.000	0.000	0.000	0.007
20	8	0.001	0.006	0.003	0.000	0.000	0.000	0.000	0.006
20	9	0.001	0.006	0.003	0.000	0.000	0.000	0.000	0.006
30	3	0.001	0.006	0.004	0.002	0.000	0.001	0.000	0.016
30	4	0.001	0.005	0.004	0.002	0.000	0.001	0.000	0.006
30	5	0.001	0.005	0.004	0.000	0.000	0.001	0.000	0.008
30	6	0.001	0.005	0.004	0.001	0.000	0.000	0.000	0.005
30	7	0.001	0.005	0.004	0.001	0.000	0.000	0.000	0.003
30	8	0.001	0.005	0.004	0.001	0.000	0.000	0.000	0.003
30	9	0.001	0.005	0.004	0.001	0.000	0.000	0.000	0.002

Table M1 (continued)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
40	3	0.001	0.005	0.004	0.001	0.000	0.001	0.000	0.016
40	4	0.001	0.005	0.004	0.001	0.000	0.001	0.000	0.008
40	5	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.006
40	6	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.004
40	7	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.003
40	8	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.002
40	9	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.001
50	3	0.001	0.005	0.004	0.001	0.000	0.001	0.000	0.015
50	4	0.001	0.004	0.004	0.001	0.000	0.001	0.000	0.003
50	5	0.001	0.003	0.004	0.000	0.000	0.000	0.000	0.004
50	6	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.003
50	7	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.001
50	8	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.001
50	9	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.000
60	3	0.001	0.005	0.004	0.001	0.000	0.001	0.000	0.013
60	4	0.001	0.004	0.004	0.001	0.000	0.001	0.000	0.005
60	5	0.001	0.003	0.004	0.000	0.000	0.000	0.000	0.003
60	6	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.002
60	7	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.000
60	8	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.000
60	9	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero. LB estimation probl.= lower bound estimation problems (ICCs ≤ -1/[k-1]). UB estimation probl. = upper bound estimation problems (ICCs ≥ 1).

Table M2

Relative Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Setting Negative ICCs to Zero Aggregated Across Replications (Negative Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	.161	.493	.709	-0.315	-0.044	0.153	0.074	0.203	0.534
14	4	.254	.549	.742	-0.360	-0.041	0.067	0.056	0.172	0.549
14	5	.349	.586	.791	-0.373	-0.039	0.025	0.043	0.152	0.552
14	6	.356	.613	.794	-0.390	-0.037	0.019	0.032	0.139	0.560
14	7	.404	.634	.804	-0.399	-0.035	0.015	0.025	0.128	0.563
14	8	.423	.651	.813	-0.386	-0.033	0.013	0.017	0.120	0.554
14	9	.450	.662	.829	-0.399	-0.031	0.011	0.011	0.114	0.560
20	3	.268	.541	.753	-0.251	-0.037	0.153	0.073	0.189	0.485
20	4	.340	.605	.777	-0.287	-0.033	0.067	0.053	0.156	0.498
20	5	.413	.649	.815	-0.282	-0.030	0.022	0.039	0.136	0.491
20	6	.459	.682	.846	-0.307	-0.027	0.017	0.029	0.122	0.504
20	7	.449	.702	.849	-0.311	-0.025	0.013	0.021	0.111	0.504
20	8	.497	.722	.865	-0.302	-0.023	0.009	0.013	0.102	0.496
20	9	.495	.736	.879	-0.308	-0.022	0.007	0.007	0.095	0.500
30	3	.316	.601	.770	-0.170	-0.030	0.159	0.070	0.172	0.435
30	4	.435	.678	.844	-0.175	-0.025	0.075	0.049	0.138	0.403
30	5	.459	.730	.870	-0.180	-0.021	0.025	0.035	0.117	0.402
30	6	.562	.764	.877	-0.188	-0.018	0.014	0.024	0.102	0.404
30	7	.576	.792	.900	-0.196	-0.016	0.010	0.016	0.090	0.408
30	8	.580	.813	.911	-0.189	-0.014	0.007	0.009	0.080	0.398
30	9	.612	.830	.925	-0.194	-0.013	0.006	0.003	0.072	0.402

Table M2 (continued)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
40	3	.386	.650	.809	-0.133	-0.025	0.157	0.065	0.159	0.413
40	4	.488	.734	.865	-0.109	-0.019	0.083	0.044	0.125	0.330
40	5	.584	.789	.893	-0.103	-0.016	0.031	0.031	0.103	0.312
40	6	.607	.828	.912	-0.104	-0.013	0.014	0.021	0.088	0.308
40	7	.667	.857	.933	-0.104	-0.010	0.010	0.013	0.075	0.302
40	8	.672	.877	.948	-0.117	-0.009	0.007	0.007	0.064	0.312
40	9	.708	.896	.956	-0.110	-0.007	0.005	0.001	0.055	0.301
50	3	.409	.687	.828	-0.100	-0.021	0.152	0.059	0.150	0.396
50	4	.533	.777	.902	-0.079	-0.016	0.076	0.042	0.115	0.301
50	5	.647	.833	.911	-0.069	-0.012	0.032	0.028	0.093	0.228
50	6	.703	.874	.936	-0.060	-0.009	0.014	0.019	0.076	0.218
50	7	.717	.903	.954	-0.056	-0.007	0.009	0.012	0.063	0.212
50	8	.790	.926	.967	-0.052	-0.005	0.006	0.005	0.051	0.201
50	9	.787	.944	.980	-0.053	-0.004	0.004	0.000	0.040	0.199
60	3	.466	.715	.841	-0.085	-0.019	0.146	0.055	0.142	0.382
60	4	.599	.808	.909	-0.057	-0.013	0.075	0.039	0.107	0.282
60	5	.685	.867	.934	-0.039	-0.010	0.032	0.027	0.084	0.196
60	6	.737	.908	.954	-0.030	-0.007	0.014	0.018	0.067	0.146
60	7	.808	.937	.972	-0.029	-0.005	0.009	0.011	0.053	0.131
60	8	.863	.959	.981	-0.022	-0.003	0.005	0.005	0.040	0.126
60	9	.878	.977	.992	-0.026	-0.002	0.003	0.000	0.026	0.122

Note. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criterion, they reflect the minimum and maximum value across replications. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table M3

Absolute Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Setting Negative ICCs to Zero Aggregated Across Replications (Negative Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	0.163	0.228	0.310	.000	.368	.970	.062	.269	.506	0	0	0	0	1.684	11
14	4	0.148	0.205	0.260	.071	.471	.974	.074	.222	.455	0	0	0	0	1.026	7
14	5	0.138	0.191	0.243	.183	.537	.732	.062	.182	.386	0	0	0	0	0.708	4
14	6	0.136	0.183	0.226	.315	.741	.994	.040	.123	.250	0	0	0	0	0.087	2
14	7	0.135	0.177	0.215	.415	.617	.770	.028	.107	.210	0	0	0	0	0.458	3
14	8	0.141	0.173	0.211	.504	.739	.992	.017	.075	.153	0	0	0	0	0.055	2
14	9	0.136	0.170	0.201	.511	.664	.779	.011	.063	.131	0	0	0	0	0.338	3
20	3	0.156	0.218	0.308	.071	.543	.818	.057	.241	.455	0	0	1	0	1.389	12
20	4	0.142	0.196	0.255	.311	.620	.991	.028	.146	.318	0	0	0	0	0.812	7
20	5	0.141	0.183	0.233	.443	.667	.984	.028	.138	.335	0	0	0	0	0.523	5
20	6	0.139	0.176	0.214	.544	.700	.823	.023	.093	.216	0	0	0	0	0.374	4
20	7	0.133	0.170	0.205	.592	.722	.831	.017	.076	.170	0	0	0	0	0.289	3
20	8	0.132	0.166	0.195	.601	.739	.819	.006	.055	.125	0	0	0	0	0.226	3
20	9	0.136	0.163	0.192	.656	.753	.839	.011	.045	.114	0	0	0	0	0.191	3
30	3	0.144	0.207	0.288	.332	.735	.995	.034	.207	.455	0	0	0	0	0.831	10
30	4	0.142	0.186	0.243	.552	.809	.997	.017	.128	.335	0	0	0	0	0.161	3
30	5	0.138	0.174	0.216	.646	.765	.989	.011	.096	.267	0	0	0	0	0.312	4
30	6	0.129	0.166	0.200	.676	.795	.995	.011	.064	.159	0	0	0	0	0.143	3
30	7	0.132	0.161	0.186	.720	.816	.995	.006	.048	.131	0	0	0	0	0.048	2
30	8	0.131	0.157	0.182	.734	.820	.995	.006	.036	.085	0	0	0	0	0.032	2
30	9	0.129	0.155	0.181	.750	.828	.995	.000	.029	.068	0	0	0	0	0.020	2

Table M3 (continued)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
40	3	0.148	0.199	0.289	.515	.752	.991	.028	.182	.420	0	0	0	0	0.765	9
40	4	0.133	0.179	0.235	.624	.807	.997	.011	.110	.341	0	0	0	0	0.270	5
40	5	0.131	0.167	0.211	.712	.815	.992	.006	.072	.227	0	0	0	0	0.152	3
40	6	0.133	0.160	0.187	.759	.829	.906	.006	.048	.148	0	0	0	0	0.084	2
40	7	0.128	0.155	0.184	.783	.840	.901	.000	.034	.097	0	0	0	0	0.034	1
40	8	0.130	0.151	0.174	.796	.848	.902	.000	.026	.062	0	0	0	0	0.017	2
40	9	0.130	0.149	0.170	.811	.856	.993	.000	.020	.051	0	0	0	0	0.007	1
50	3	0.144	0.193	0.279	.605	.792	.913	.023	.163	.420	0	0	0	0	0.584	8
50	4	0.138	0.173	0.232	.715	.851	.998	.006	.091	.290	0	0	0	0	0.038	2
50	5	0.135	0.162	0.205	.773	.845	.994	.006	.057	.216	0	0	0	0	0.081	2
50	6	0.135	0.155	0.184	.802	.855	.922	.000	.037	.119	0	0	0	0	0.034	1
50	7	0.131	0.151	0.173	.821	.865	.994	.000	.026	.085	0	0	0	0	0.007	2
50	8	0.131	0.147	0.167	.835	.871	.905	.000	.019	.057	0	0	0	0	0.002	1
50	9	0.130	0.145	0.158	.849	.877	.904	.000	.016	.040	0	0	0	0	0.001	1
60	3	0.140	0.189	0.271	.645	.822	.994	.017	.150	.415	0	0	0	0	0.452	7
60	4	0.138	0.169	0.223	.762	.851	.997	.006	.080	.284	0	0	0	0	0.148	3
60	5	0.134	0.159	0.198	.808	.863	.995	.000	.047	.188	0	0	0	0	0.057	2
60	6	0.134	0.152	0.179	.833	.873	.995	.000	.030	.102	0	0	0	0	0.015	2
60	7	0.132	0.148	0.166	.851	.881	.923	.000	.020	.068	0	0	0	0	0.001	1
60	8	0.131	0.144	0.160	.867	.887	.914	.000	.015	.045	0	0	0	0	0	1
60	9	0.131	0.142	0.151	.875	.893	.910	.006	.014	.028	0	0	0	0	0	0
70	9	-	0.140	-	-	.904	-	-	.011	-	-	0	-	-	0	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 9 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The minimum and maximum reflect the minimum and maximum value across replications. The reliability calculation required Fisher's *Z*-transformed ICCs; thus, participants affected by estimation problems were excluded from these analyses. The minimum sample size for the reliability calculation ranged from 164 to 176 across conditions. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. LB = lower bound estimation problems (ICCs ≤ -1/[k-1]). UB = upper bound estimation problems (ICCs ≥ 1). N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table M4

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Setting Negative ICCs to Zero Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Positive Emotion Differentiation in the Replication Data Set)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
14	3	0.001	0.005	0.003	0.003	0.000	0.001	0.000	0.004
14	4	0.001	0.004	0.003	0.002	0.000	0.000	0.000	0.002
14	5	0.000	0.004	0.003	0.001	0.000	0.000	0.001	0.000
20	3	0.001	0.006	0.003	0.003	0.000	0.001	0.000	0.003
20	4	0.001	0.004	0.003	0.001	0.000	0.000	0.000	0.001
20	5	0.000	0.003	0.003	0.001	0.000	0.000	0.000	0.000
30	3	0.001	0.006	0.003	0.002	0.000	0.001	0.000	0.001
30	4	0.001	0.004	0.003	0.001	0.000	0.000	0.000	0.000
30	5	0.000	0.002	0.002	0.000	0.000	0.000	0.000	0.000
40	3	0.001	0.006	0.003	0.002	0.000	0.001	0.000	0.001
40	4	0.001	0.004	0.003	0.001	0.000	0.000	0.000	0.000
40	5	0.000	0.001	0.002	0.000	0.000	0.000	0.000	0.000
50	3	0.001	0.006	0.003	0.001	0.000	0.001	0.000	0.000
50	4	0.001	0.004	0.003	0.001	0.000	0.000	0.000	0.000
50	5	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000
60	3	0.001	0.006	0.003	0.001	0.000	0.001	0.000	0.000
60	4	0.001	0.004	0.003	0.000	0.000	0.000	0.000	0.000
60	5	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero. LB estimation probl.= lower bound estimation problems (ICCs ≤ -1/[k-1]). UB estimation probl. = upper bound estimation problems (ICCs ≥ 1).

Table M5

Relative Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Setting Negative ICCs to Zero Aggregated Across Replications (Positive Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	.338	.574	.735	-0.061	-0.011	0.101	0.112	0.185	0.394
14	4	.421	.665	.799	-0.082	-0.011	0.052	0.080	0.144	0.309
14	5	.596	.725	.823	-0.099	-0.009	0.031	0.055	0.113	0.267
20	3	.398	.639	.781	-0.049	-0.009	0.091	0.096	0.168	0.401
20	4	.520	.739	.846	-0.051	-0.008	0.046	0.067	0.124	0.301
20	5	.702	.804	.877	-0.067	-0.006	0.023	0.044	0.090	0.209
30	3	.473	.708	.849	-0.045	-0.008	0.078	0.081	0.152	0.407
30	4	.635	.815	.900	-0.026	-0.006	0.042	0.053	0.105	0.294
30	5	.799	.882	.926	-0.032	-0.003	0.017	0.031	0.066	0.141
40	3	.512	.751	.861	-0.045	-0.008	0.072	0.074	0.143	0.412
40	4	.698	.862	.924	-0.018	-0.004	0.038	0.044	0.094	0.289
40	5	.872	.929	.960	-0.016	-0.002	0.012	0.022	0.049	0.096
50	3	.572	.780	.876	-0.045	-0.007	0.070	0.067	0.136	0.415
50	4	.748	.894	.941	-0.017	-0.004	0.036	0.036	0.086	0.286
50	5	.923	.961	.976	-0.008	-0.001	0.008	0.015	0.036	0.068
60	3	.602	.802	.883	-0.044	-0.007	0.070	0.056	0.132	0.418
60	4	.785	.918	.957	-0.016	-0.003	0.035	0.032	0.080	0.283
60	5	.967	.983	.990	-0.003	0.000	0.004	0.008	0.023	0.041

Note. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criterion, they reflect the minimum and maximum value across replications. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table M6

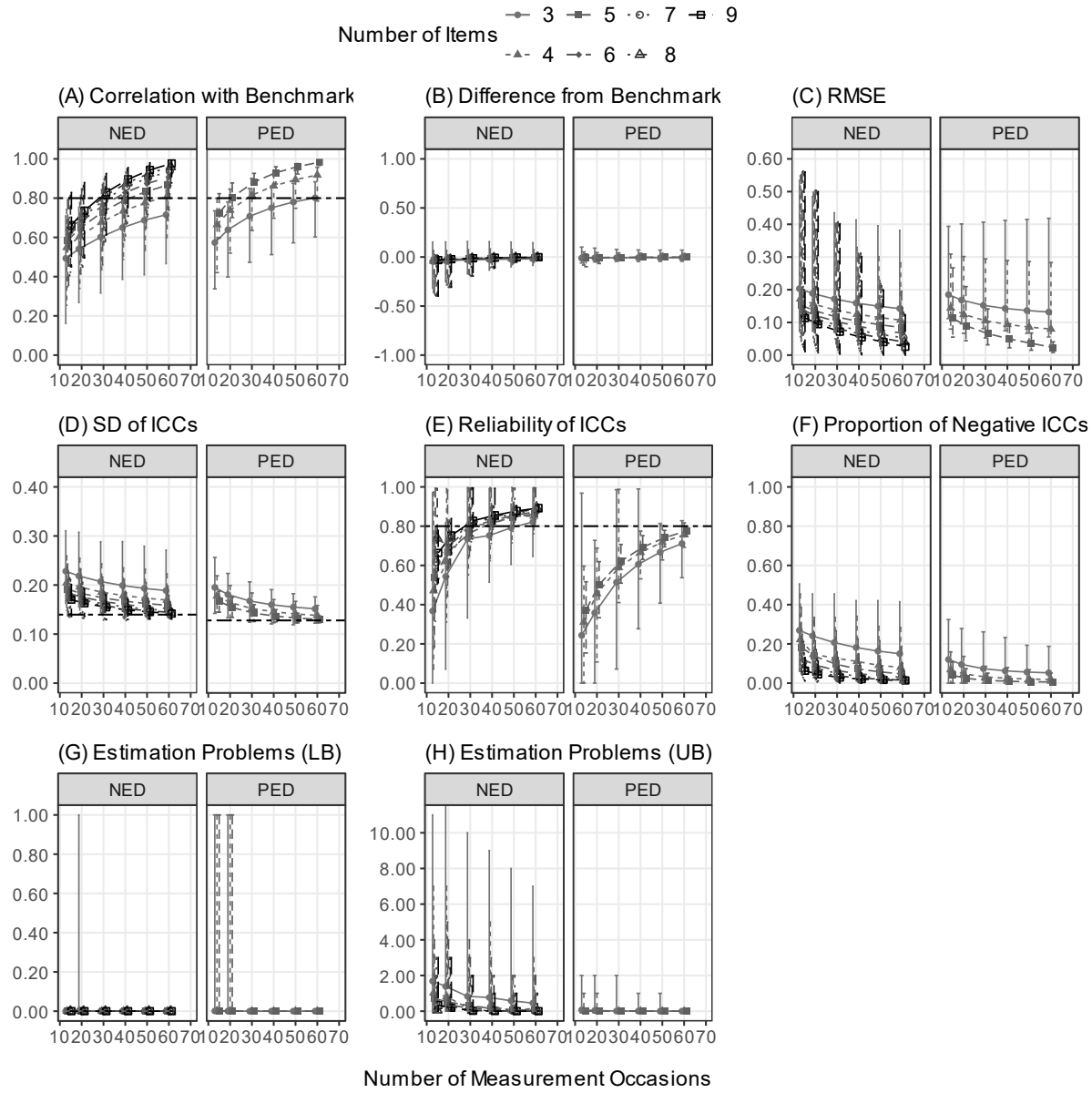
Absolute Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Setting Negative ICCs to Zero Aggregated Across Replications (Positive Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	0.142	0.195	0.256	.000	.244	.968	.000	.121	.324	0	0.001	1	0	0.068	2
14	4	0.145	0.178	0.219	.000	.311	.596	.000	.069	.159	0	0.001	1	0	0.014	1
14	5	0.145	0.167	0.187	.154	.373	.517	.000	.04	.108	0	0.002	1	0	0	0
20	3	0.133	0.181	0.224	.000	.358	.728	.000	.096	.278	0	0	1	0	0.038	2
20	4	0.128	0.164	0.199	.108	.457	.688	.000	.047	.136	0	0	1	0	0.005	1
20	5	0.134	0.154	0.175	.333	.502	.619	.000	.024	.074	0	0.001	1	0	0	0
30	3	0.124	0.168	0.206	.072	.514	.986	.000	.074	.261	0	0	0	0	0.009	2
30	4	0.127	0.152	0.184	.411	.589	.988	.000	.032	.085	0	0	0	0	0	0
30	5	0.126	0.143	0.160	.510	.621	.707	.000	.014	.051	0	0	0	0	0	0
40	3	0.121	0.160	0.191	.277	.605	.990	.000	.063	.233	0	0	0	0	0.003	1
40	4	0.123	0.145	0.170	.532	.667	.775	.000	.025	.080	0	0	0	0	0	0
40	5	0.125	0.136	0.150	.631	.693	.752	.000	.009	.034	0	0	0	0	0	0
50	3	0.119	0.155	0.182	.408	.667	.814	.000	.056	.193	0	0	0	0	0	1
50	4	0.123	0.141	0.167	.628	.718	.804	.000	.021	.068	0	0	0	0	0	0
50	5	0.123	0.132	0.142	.695	.741	.779	.000	.007	.023	0	0	0	0	0	0
60	3	0.122	0.152	0.176	.537	.711	.828	.000	.053	.188	0	0	0	0	0	0
60	4	0.123	0.138	0.158	.692	.756	.817	.000	.019	.051	0	0	0	0	0	0
60	5	0.123	0.130	0.137	.750	.776	.800	.000	.006	.017	0	0	0	0	0	0
70	5	-	0.128	-	-	.803	-	-	.006	-	-	0	-	-	0	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 5 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The minimum and maximum reflect the minimum and maximum value across replications. The reliability calculation required Fisher's *Z*-transformed ICCs; thus, participants affected by estimation problems were excluded from these analyses. The minimum sample size for the reliability calculation ranged from 174 to 176 across conditions. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. LB = lower bound estimation problems (ICCs ≤ -1/[k-1]). UB = upper bound estimation problems (ICCs ≥ 1). N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Figure M1

Results of the Evaluation Criteria for Raw Intraclass Correlations (ICCs) for Each Condition When Setting Negative ICCs to Zero



Note. Panels (A) to (H) depict the evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (H) are absolute criteria. The panels present results for negative emotion differentiation (NED; left) and positive emotion differentiation (PED; right). The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across replications. The plots display mean values for each item number condition. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E), and the standard deviation (*SD*) of ICCs in the benchmark condition in (D). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs. LB = lower bound. UB = upper bound.

Table M7

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Excluding Negative ICCs Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Negative Emotion Differentiation in the Replication Data Set)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
14	3	0.001	0.007	0.006	0.002	0.000	0.001	0.000	0.022
14	4	0.001	0.007	0.005	0.002	0.000	0.001	0.000	0.014
14	5	0.001	0.007	0.005	0.001	0.000	0.001	0.000	0.011
14	6	0.001	0.006	0.004	0.003	0.000	0.000	0.000	0.004
14	7	0.001	0.006	0.004	0.001	0.000	0.000	0.000	0.009
14	8	0.001	0.006	0.003	0.002	0.000	0.000	0.000	0.003
14	9	0.001	0.006	0.003	0.001	0.000	0.000	0.000	0.008
20	3	0.001	0.007	0.005	0.002	0.000	0.001	0.000	0.021
20	4	0.001	0.006	0.004	0.001	0.000	0.001	0.000	0.013
20	5	0.001	0.006	0.005	0.001	0.000	0.001	0.000	0.010
20	6	0.001	0.006	0.004	0.001	0.000	0.000	0.000	0.008
20	7	0.001	0.006	0.003	0.001	0.000	0.000	0.000	0.007
20	8	0.001	0.006	0.003	0.001	0.000	0.000	0.000	0.006
20	9	0.001	0.006	0.003	0.000	0.000	0.000	0.000	0.006
30	3	0.001	0.006	0.005	0.002	0.000	0.001	0.000	0.016
30	4	0.001	0.005	0.004	0.002	0.000	0.001	0.000	0.006
30	5	0.001	0.005	0.004	0.001	0.000	0.001	0.000	0.008
30	6	0.001	0.005	0.004	0.001	0.000	0.000	0.000	0.005
30	7	0.001	0.005	0.004	0.001	0.000	0.000	0.000	0.003
30	8	0.001	0.005	0.004	0.001	0.000	0.000	0.000	0.002
30	9	0.001	0.005	0.004	0.001	0.000	0.000	0.000	0.002

Table M7 (continued)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
40	3	0.001	0.005	0.005	0.001	0.000	0.001	0.000	0.016
40	4	0.001	0.005	0.004	0.001	0.000	0.001	0.000	0.008
40	5	0.001	0.004	0.004	0.001	0.000	0.000	0.000	0.006
40	6	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.004
40	7	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.003
40	8	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.002
40	9	0.001	0.004	0.004	0.000	0.000	0.000	0.000	0.001
50	3	0.001	0.005	0.005	0.001	0.000	0.001	0.000	0.014
50	4	0.001	0.004	0.004	0.001	0.000	0.001	0.000	0.003
50	5	0.001	0.003	0.004	0.001	0.000	0.000	0.000	0.004
50	6	0.001	0.003	0.004	0.000	0.000	0.000	0.000	0.003
50	7	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.001
50	8	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.001
50	9	0.000	0.003	0.004	0.000	0.000	0.000	0.000	0.000
60	3	0.001	0.006	0.006	0.001	0.000	0.001	0.000	0.012
60	4	0.001	0.004	0.005	0.001	0.000	0.001	0.000	0.006
60	5	0.001	0.003	0.005	0.000	0.000	0.000	0.000	0.003
60	6	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.002
60	7	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.001
60	8	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.000
60	9	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero. LB estimation probl.= lower bound estimation problems (ICCs $\leq$ -1/[k-1]). UB estimation probl. = upper bound estimation problems (ICCs $\geq$ 1).

Table M8

Relative Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Excluding Negative ICCs Aggregated Across Replications (Negative Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	.134	.478	.739	-0.191	0.011	0.233	0.115	0.204	0.504
14	4	.191	.542	.785	-0.252	0.002	0.270	0.079	0.167	0.477
14	5	.299	.592	.794	-0.269	-0.006	0.172	0.065	0.145	0.480
14	6	.278	.603	.802	-0.333	-0.020	0.048	0.051	0.135	0.522
14	7	.378	.641	.814	-0.376	-0.020	0.057	0.041	0.124	0.546
14	8	.406	.642	.821	-0.384	-0.025	0.026	0.032	0.118	0.551
14	9	.350	.653	.835	-0.393	-0.025	0.024	0.026	0.112	0.555
20	3	.196	.525	.757	-0.140	0.011	0.234	0.106	0.187	0.487
20	4	.277	.586	.787	-0.226	-0.011	0.093	0.073	0.153	0.455
20	5	.345	.651	.839	-0.220	-0.008	0.140	0.054	0.131	0.442
20	6	.404	.674	.848	-0.269	-0.016	0.053	0.041	0.119	0.475
20	7	.483	.708	.852	-0.292	-0.016	0.052	0.033	0.108	0.491
20	8	.455	.719	.870	-0.299	-0.018	0.034	0.026	0.100	0.494
20	9	.478	.733	.878	-0.309	-0.018	0.035	0.021	0.093	0.499
30	3	.284	.579	.780	-0.086	0.009	0.205	0.092	0.170	0.450
30	4	.363	.663	.832	-0.131	-0.006	0.132	0.065	0.134	0.374
30	5	.480	.724	.858	-0.152	-0.008	0.102	0.046	0.114	0.372
30	6	.423	.759	.887	-0.170	-0.012	0.057	0.034	0.100	0.386
30	7	.580	.791	.903	-0.185	-0.011	0.042	0.025	0.089	0.398
30	8	.576	.810	.917	-0.186	-0.011	0.031	0.019	0.080	0.397
30	9	.627	.827	.921	-0.188	-0.011	0.032	0.016	0.071	0.396

Table M8 (continued)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
40	3	.306	.624	.805	-0.063	0.007	0.181	0.080	0.157	0.419
40	4	.361	.718	.858	-0.071	-0.004	0.103	0.056	0.123	0.333
40	5	.565	.780	.883	-0.085	-0.007	0.088	0.041	0.102	0.291
40	6	.560	.821	.908	-0.097	-0.009	0.058	0.029	0.087	0.296
40	7	.655	.853	.930	-0.098	-0.008	0.039	0.021	0.075	0.296
40	8	.674	.876	.950	-0.109	-0.007	0.025	0.016	0.064	0.302
40	9	.708	.892	.956	-0.112	-0.007	0.025	0.012	0.055	0.305
50	3	.348	.658	.831	-0.063	0.006	0.171	0.072	0.148	0.400
50	4	.520	.756	.877	-0.047	-0.005	0.090	0.051	0.114	0.301
50	5	.603	.823	.923	-0.047	-0.006	0.084	0.037	0.092	0.223
50	6	.666	.867	.935	-0.050	-0.007	0.054	0.026	0.076	0.214
50	7	.745	.899	.951	-0.055	-0.006	0.031	0.018	0.063	0.218
50	8	.739	.923	.964	-0.055	-0.005	0.017	0.013	0.052	0.210
50	9	.794	.941	.977	-0.054	-0.004	0.017	0.009	0.040	0.204
60	3	.407	.686	.870	-0.063	0.005	0.157	0.066	0.141	0.383
60	4	.518	.790	.914	-0.044	-0.004	0.117	0.046	0.106	0.279
60	5	.661	.857	.928	-0.033	-0.006	0.080	0.034	0.084	0.193
60	6	.715	.902	.951	-0.026	-0.005	0.045	0.024	0.067	0.145
60	7	.804	.934	.970	-0.024	-0.004	0.024	0.016	0.053	0.135
60	8	.834	.957	.980	-0.022	-0.003	0.011	0.011	0.040	0.125
60	9	.848	.976	.994	-0.022	-0.002	0.008	0.006	0.026	0.118

Note. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criterion, they reflect the minimum and maximum value across replications. Since participants with negative ICCs were excluded, the minimum sample size ranged from 88 to 171 across conditions. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table M9

Absolute Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Excluding Negative ICCs Aggregated Across Replications (Negative Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	0.158	0.229	0.340	.000	.385	.982	.074	.268	.494	0	0.001	1	0	1.682	12
14	4	0.143	0.198	0.285	.000	.421	.984	.085	.223	.455	0	0	0	0	1.017	7
14	5	0.132	0.184	0.237	.024	.467	.730	.062	.183	.409	0	0	0	0	0.722	6
14	6	0.132	0.177	0.224	.231	.711	.993	.034	.123	.261	0	0	1	0	0.082	2
14	7	0.129	0.171	0.208	.329	.560	.750	.034	.108	.227	0	0	0	0	0.451	3
14	8	0.126	0.169	0.206	.382	.714	.994	.011	.075	.159	0	0	0	0	0.056	2
14	9	0.133	0.166	0.198	.442	.624	.752	.011	.063	.136	0	0	0	0	0.312	3
20	3	0.147	0.215	0.334	.000	.536	.982	.062	.240	.460	0	0	0	0	1.396	12
20	4	0.141	0.191	0.268	.273	.593	.986	.028	.147	.330	0	0	0	0	0.816	6
20	5	0.125	0.176	0.234	.312	.621	.987	.028	.138	.375	0	0	0	0	0.533	4
20	6	0.129	0.169	0.209	.481	.661	.816	.011	.093	.210	0	0	0	0	0.367	3
20	7	0.130	0.164	0.198	.521	.685	.823	.017	.076	.176	0	0	0	0	0.285	3
20	8	0.132	0.161	0.192	.575	.711	.818	.011	.054	.125	0	0	0	0	0.228	3
20	9	0.132	0.159	0.192	.603	.728	.824	.006	.045	.102	0	0	0	0	0.191	3
30	3	0.145	0.201	0.313	.292	.717	.997	.045	.207	.438	0	0	0	0	0.836	9
30	4	0.131	0.179	0.251	.454	.790	.998	.023	.128	.335	0	0	0	0	0.170	4
30	5	0.133	0.167	0.216	.545	.735	.890	.011	.096	.290	0	0	0	0	0.313	4
30	6	0.126	0.161	0.195	.618	.774	.995	.006	.064	.170	0	0	0	0	0.134	3
30	7	0.126	0.157	0.186	.658	.795	.995	.006	.048	.131	0	0	0	0	0.044	1
30	8	0.124	0.154	0.180	.696	.804	.995	.000	.036	.085	0	0	0	0	0.031	2
30	9	0.127	0.152	0.178	.726	.814	.995	.000	.029	.068	0	0	0	0	0.022	1

Table M9 (continued)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
40	3	0.137	0.193	0.307	.309	.738	.994	.028	.182	.432	0	0	0	0	0.742	8
40	4	0.128	0.172	0.242	.568	.789	.997	.017	.110	.324	0	0	0	0	0.275	5
40	5	0.128	0.161	0.205	.666	.796	.994	.011	.072	.250	0	0	0	0	0.155	3
40	6	0.127	0.155	0.191	.713	.812	.992	.000	.048	.131	0	0	0	0	0.087	2
40	7	0.127	0.151	0.177	.757	.826	.905	.000	.034	.097	0	0	0	0	0.041	2
40	8	0.126	0.149	0.172	.766	.837	.887	.000	.026	.074	0	0	0	0	0.020	1
40	9	0.126	0.146	0.168	.790	.845	.993	.000	.021	.057	0	0	0	0	0.008	1
50	3	0.134	0.187	0.298	.533	.778	.925	.011	.163	.403	0	0	0	0	0.575	9
50	4	0.129	0.167	0.235	.665	.836	.998	.006	.091	.295	0	0	0	0	0.041	3
50	5	0.127	0.157	0.204	.735	.832	.997	.000	.057	.205	0	0	0	0	0.079	2
50	6	0.129	0.152	0.184	.778	.843	.916	.000	.037	.108	0	0	0	0	0.038	2
50	7	0.127	0.148	0.173	.803	.855	.994	.000	.026	.080	0	0	0	0	0.006	1
50	8	0.128	0.145	0.167	.822	.863	.905	.000	.019	.045	0	0	0	0	0.004	1
50	9	0.126	0.143	0.158	.832	.869	.900	.000	.016	.040	0	0	0	0	0	1
60	3	0.136	0.182	0.286	.609	.810	.996	.017	.150	.409	0	0	0	0	0.447	7
60	4	0.130	0.164	0.231	.725	.838	.998	.006	.080	.295	0	0	0	0	0.146	3
60	5	0.129	0.154	0.197	.782	.853	.995	.006	.047	.170	0	0	0	0	0.057	2
60	6	0.130	0.149	0.171	.817	.864	.995	.000	.030	.091	0	0	0	0	0.013	1
60	7	0.130	0.145	0.165	.841	.874	.926	.000	.020	.068	0	0	0	0	0.002	1
60	8	0.129	0.142	0.160	.854	.881	.917	.000	.015	.045	0	0	0	0	0	1
60	9	0.126	0.140	0.149	.861	.886	.905	.006	.014	.028	0	0	0	0	0	0
70	9	-	0.138	-	-	.900	-	-	.011	-	-	0	-	-	0	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 9 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The minimum and maximum reflect the minimum and maximum value across replications. Since participants with negative ICCs (and/or, for reliability analysis, affected by estimation problems) were excluded, the minimum sample size ranged from 88 to 171 across conditions. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. LB = lower bound estimation problems (ICCs ≤ -1/[k-1]). UB = upper bound estimation problems (ICCs ≥ 1). N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table M10

Monte Carlo Standard Error (MCSE) for the Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Excluding Negative ICCs Separate for All Simulation Conditions Using Random Draws of Measurement Occasions (Positive Emotion Differentiation in the Replication Data Set)

Condition		MCSE							
N_{occ}	N_{items}	Correlation	Difference	RMSE	Reliability	*SD*	% Neg ICC	LB estimation probl.	UB estimation probl.
14	3	0.001	0.006	0.004	0.003	0.000	0.001	0.001	0.004
14	4	0.001	0.005	0.003	0.002	0.000	0.000	0.000	0.002
14	5	0.001	0.003	0.003	0.001	0.000	0.000	0.001	0.000
20	3	0.001	0.006	0.003	0.003	0.000	0.001	0.000	0.003
20	4	0.001	0.004	0.003	0.002	0.000	0.000	0.000	0.001
20	5	0.000	0.003	0.002	0.001	0.000	0.000	0.000	0.000
30	3	0.001	0.006	0.003	0.003	0.000	0.001	0.000	0.002
30	4	0.001	0.004	0.003	0.001	0.000	0.000	0.000	0.000
30	5	0.000	0.002	0.002	0.000	0.000	0.000	0.000	0.000
40	3	0.001	0.007	0.003	0.002	0.000	0.001	0.000	0.001
40	4	0.001	0.004	0.003	0.001	0.000	0.000	0.000	0.000
40	5	0.000	0.001	0.002	0.000	0.000	0.000	0.000	0.000
50	3	0.001	0.007	0.003	0.002	0.000	0.001	0.000	0.000
50	4	0.001	0.004	0.003	0.001	0.000	0.000	0.000	0.000
50	5	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000
60	3	0.001	0.006	0.003	0.002	0.000	0.001	0.000	0.000
60	4	0.001	0.004	0.003	0.001	0.000	0.000	0.000	0.000
60	5	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000

Note. Since the difference ("bias") and RMSE was calculated across replications per participant, we calculated the MCSE per participant according to the formulas in Table 3 in Siepe et al. (2024). For the difference and the RMSE, we report the maximum person-level MCSE as most conservative measure. For the remaining evaluation criteria, which were averaged across replications, the MCSE was calculated using the formula for the mean of generic statistic *G*, according to Siepe et al. (2024). N_{occ} = number of measurement occasions. N_{items} = number of emotion items. RMSE = root mean squared error. *SD* = standard deviation. % Neg ICC = proportion of participants with an ICC below zero. LB estimation probl.= lower bound estimation problems (ICCs ≤ -1/[k-1]). UB estimation probl. = upper bound estimation problems (ICCs ≥ 1).

Table M11

Relative Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Excluding Negative ICCs Aggregated Across Replications (Positive Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference			RMSE		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	.301	.534	.708	-0.050	0.020	0.295	0.117	0.180	0.440
14	4	.405	.636	.773	-0.054	0.002	0.119	0.089	0.140	0.315
14	5	.540	.705	.811	-0.079	-0.004	0.065	0.059	0.111	0.237
20	3	.362	.608	.766	-0.042	0.013	0.308	0.103	0.165	0.453
20	4	.539	.717	.841	-0.034	-0.001	0.114	0.069	0.123	0.304
20	5	.657	.791	.878	-0.052	-0.004	0.047	0.045	0.089	0.189
30	3	.455	.682	.826	-0.042	0.007	0.326	0.091	0.151	0.473
30	4	.596	.800	.884	-0.022	-0.002	0.115	0.053	0.105	0.295
30	5	.798	.875	.921	-0.032	-0.003	0.035	0.032	0.066	0.139
40	3	.510	.732	.837	-0.043	0.005	0.331	0.075	0.142	0.485
40	4	.700	.851	.919	-0.017	-0.002	0.116	0.043	0.094	0.290
40	5	.864	.926	.953	-0.015	-0.002	0.024	0.024	0.049	0.098
50	3	.538	.766	.869	-0.044	0.004	0.363	0.064	0.136	0.500
50	4	.744	.887	.942	-0.016	-0.002	0.124	0.035	0.086	0.287
50	5	.921	.959	.973	-0.010	-0.001	0.016	0.017	0.036	0.068
60	3	.597	.791	.874	-0.044	0.004	0.422	0.055	0.132	0.525
60	4	.784	.913	.956	-0.016	-0.002	0.141	0.029	0.080	0.284
60	5	.959	.982	.990	-0.004	0.000	0.009	0.011	0.023	0.042

Note. For the difference in ICCs between the simulation conditions and the benchmark and for the RMSE, the minimum and maximum values reflect the person-level minimum and maximum. For the remaining criterion, they reflect the minimum and maximum value across replications. Since participants with negative ICCs were excluded, the minimum sample size ranged from 119 to 172 across conditions. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Table M12

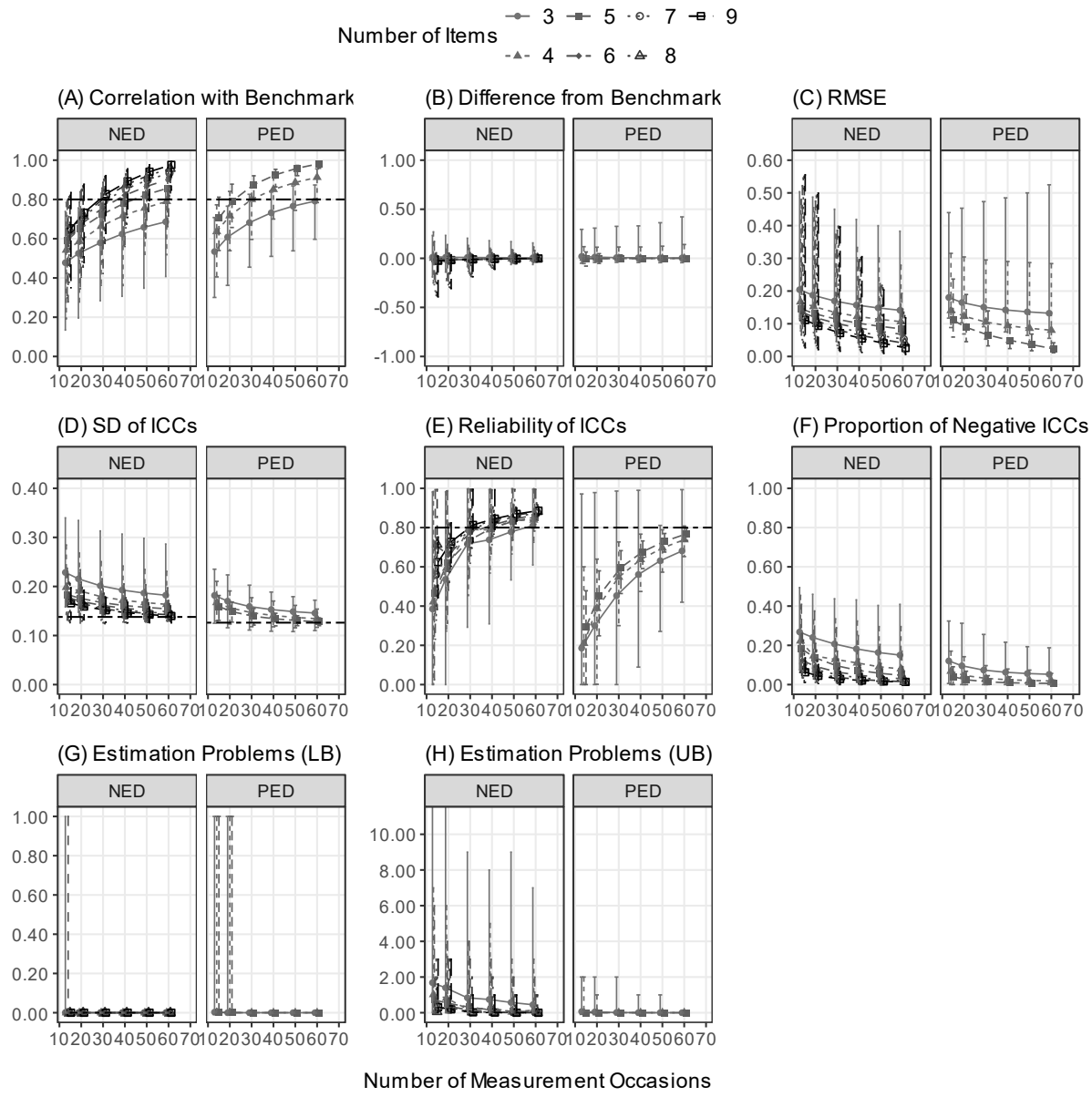
Absolute Evaluation Criteria for Raw Intraclass Correlations (ICCs) When Excluding Negative ICCs Aggregated Across Replications (Positive Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{items}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
14	3	0.125	0.182	0.235	.000	.185	.971	.006	.121	.324	0	0.003	1	0	0.067	2
14	4	0.130	0.167	0.210	.000	.210	.600	.000	.068	.170	0	0.001	1	0	0.013	2
14	5	0.131	0.159	0.182	.000	.292	.478	.000	.04	.091	0	0.001	1	0	0	0
20	3	0.116	0.170	0.224	.000	.300	.978	.000	.096	.312	0	0	1	0	0.042	2
20	4	0.123	0.156	0.191	.001	.388	.639	.000	.047	.142	0	0.001	1	0	0.004	1
20	5	0.127	0.149	0.168	.249	.453	.580	.000	.025	.068	0	0	1	0	0	0
30	3	0.111	0.159	0.203	.000	.453	.986	.000	.074	.256	0	0	0	0	0.013	2
30	4	0.118	0.147	0.177	.300	.547	.727	.000	.032	.097	0	0	0	0	0	0
30	5	0.121	0.139	0.155	.464	.596	.682	.000	.014	.045	0	0	0	0	0	0
40	3	0.109	0.153	0.189	.089	.561	.990	.000	.063	.216	0	0	0	0	0.004	1
40	4	0.118	0.141	0.170	.474	.637	.766	.000	.025	.080	0	0	0	0	0	0
40	5	0.120	0.134	0.145	.591	.677	.732	.000	.009	.040	0	0	0	0	0	0
50	3	0.108	0.149	0.179	.271	.631	.811	.000	.056	.193	0	0	0	0	0	1
50	4	0.116	0.137	0.161	.573	.696	.789	.000	.021	.074	0	0	0	0	0	0
50	5	0.120	0.130	0.140	.675	.730	.770	.000	.007	.023	0	0	0	0	0	0
60	3	0.110	0.146	0.172	.420	.681	.993	.000	.053	.188	0	0	0	0	0	0
60	4	0.117	0.134	0.154	.652	.738	.808	.000	.018	.057	0	0	0	0	0	0
60	5	0.120	0.128	0.136	.734	.768	.797	.000	.006	.023	0	0	0	0	0	0
70	5	-	0.126	-	-	.796	-	-	.006	-	-	0	-	-	0	-

Note. For comparison purposes, the last row reports the benchmark condition (70 measurement occasions, 5 emotion items, measurement occasions drawn by temporal order); thus, only absolute evaluation criteria apply. The minimum and maximum reflect the minimum and maximum value across replications. Since participants with negative ICCs (and/or, for reliability analysis, affected by estimation problems) were excluded, the minimum sample size ranged from 119 to 172 across conditions. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. LB = lower bound estimation problems (ICCs ≤ -1/[k-1]). UB = upper bound estimation problems (ICCs ≥ 1). N_{occ} = number of measurement occasions. N_{items} = number of items. Min = minimum. *M* = mean. Max = maximum.

Figure M2

Results of the Evaluation Criteria for Raw Intraclass Correlations (ICCs) for Each Condition When Excluding Negative ICCs



Note. Panels (A) to (H) depict the evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (H) are absolute criteria. The panels present results for negative emotion differentiation (NED; left) and positive emotion differentiation (PED; right). The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across replications. The plots display mean values for each item number condition. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E), and the standard deviation (*SD*) of ICCs in the benchmark condition in (D). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs. LB = lower bound. UB = upper bound.

N: Sensitivity Analysis Comparing Item Sets with Lowest and Highest Endorsement

To compare item sets with the lowest and the highest endorsement, we identified the item sets with the highest and the lowest mean value in the benchmark sample.

The sample means of the full item sets were 1.48 for negative emotion differentiation (9 items) and 2.59 for positive emotion differentiation (5 items).

For negative emotion differentiation, the 3-item set with the lowest mean value consisted of ashamed, envious, and resentful (M = 1.18), and the 3-item set with the highest mean value included insecure, lonely, and sad (M = 1.73). The 6-item set with the lowest mean value included angry, anxious, ashamed, envious, excluded, and resentful (M = 1.36), and the 6-item set with the highest mean value consisted of angry, anxious, excluded, insecure, lonely, and sad (M = 1.63). The 8-item set with the lowest mean value included angry, anxious, ashamed, envious, excluded, insecure, resentful, and sad (M = 1.44), and the 8-item set with the highest mean value consisted of angry, anxious, ashamed, envious, excluded, insecure, lonely, sad (M = 1.52).

For positive emotion differentiation, the 3-item set with the lowest mean value consisted of enthusiastic, proud, and superior (M = 2.13), and the 3-item set with the highest mean value included enthusiastic, relaxed, and successful (M = 3.03). The 4-item set with the lowest mean value included enthusiastic, proud, successful, and superior (M = 2.29), and the 4-item set with the highest mean value included enthusiastic, proud, relaxed, and successful (M = 2.85).

Table N1

Relative Evaluation Criteria When Only Comparing the Item Sets with the Lowest and the Highest Endorsement (Negative Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference			RMSE		
N_{occ}	End.	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 3 items										
14	low	.305	.422	.518	-0.458	-0.079	0.842	0.010	0.242	0.842
20	low	.344	.437	.514	-0.458	-0.074	0.842	0.009	0.232	0.842
30	low	.371	.453	.535	-0.458	-0.063	0.842	0.008	0.218	0.842
40	low	.381	.475	.535	-0.458	-0.060	0.842	0.006	0.207	0.842
50	low	.438	.491	.545	-0.458	-0.056	0.842	0.005	0.198	0.842
60	low	.426	.502	.529	-0.458	-0.052	0.842	0.005	0.189	0.842
14	high	.305	.501	.646	-0.417	0.055	0.379	0.015	0.229	0.572
20	high	.369	.537	.658	-0.413	0.067	0.371	0.013	0.209	0.551
30	high	.485	.607	.692	-0.301	0.084	0.444	0.015	0.186	0.547
40	high	.563	.646	.728	-0.277	0.092	0.416	0.014	0.171	0.509
50	high	.641	.701	.749	-0.165	0.098	0.473	0.014	0.156	0.521
60	high	.648	.721	.749	-0.158	0.101	0.535	0.012	0.145	0.538
Item sets with 6 items										
14	low	.461	.612	.743	-0.432	-0.094	0.069	0.014	0.155	0.623
20	low	.578	.677	.793	-0.313	-0.084	0.048	0.011	0.138	0.495
30	low	.639	.747	.838	-0.28	-0.075	0.052	0.010	0.118	0.501
40	low	.678	.801	.873	-0.236	-0.071	0.058	0.008	0.105	0.383
50	low	.777	.862	.897	-0.254	-0.065	0.113	0.008	0.091	0.270
60	low	.812	.893	.912	-0.234	-0.064	0.084	0.005	0.082	0.241
14	high	.420	.595	.716	-0.430	0.027	0.198	0.023	0.153	0.556
20	high	.529	.664	.786	-0.387	0.036	0.191	0.023	0.133	0.510
30	high	.632	.760	.855	-0.283	0.048	0.189	0.021	0.114	0.427
40	high	.708	.821	.893	-0.156	0.053	0.189	0.013	0.099	0.291
50	high	.771	.865	.913	-0.154	0.057	0.193	0.013	0.089	0.244
60	high	.837	.903	.923	-0.101	0.060	0.197	0.010	0.079	0.203

Table N1 (continued)

Condition		Correlation			Difference			RMSE		
N_{occ}	End.	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 8 items										
14	low	.438	.645	.825	-0.394	-0.039	0.064	0.026	0.122	0.558
20	low	.523	.718	.832	-0.311	-0.028	0.075	0.021	0.102	0.497
30	low	.641	.805	.895	-0.219	-0.019	0.064	0.014	0.080	0.411
40	low	.719	.874	.933	-0.136	-0.013	0.066	0.011	0.063	0.310
50	low	.795	.925	.960	-0.093	-0.010	0.071	0.009	0.049	0.200
60	low	.872	.959	.976	-0.083	-0.007	0.067	0.008	0.036	0.126
14	high	.478	.650	.822	-0.425	-0.018	0.058	0.024	0.120	0.569
20	high	.525	.727	.860	-0.300	-0.008	0.063	0.019	0.101	0.489
30	high	.645	.816	.922	-0.210	0.002	0.062	0.013	0.079	0.406
40	high	.705	.883	.936	-0.126	0.008	0.064	0.009	0.062	0.302
50	high	.821	.931	.963	-0.081	0.011	0.066	0.006	0.049	0.191
60	high	.876	.964	.981	-0.057	0.013	0.065	0.004	0.037	0.133

Note. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 59 to 60 replications (3-item and 6-item sets) or 555 to 556 replications (8-item sets), respectively, per occasion-number condition. The mean values were *M* = 1.18 for the low endorsement 3-item set, *M* = 1.73 for the high endorsement 3-item set, *M* = 1.36 for the low endorsement 6-item set, *M* = 1.63 for the high endorsement 6-item set, *M* = 1.44 for the low endorsement 8-item set, and *M* = 1.52 for the high endorsement 8-item set. Since the 9-item conditions included the full item set, we could not compare low and high endorsement item sets for this number of items. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. N_{occ} = number of measurement occasions. End. = endorsement (sample mean value of item set). Min = minimum. *M* = mean. Max = maximum. End. = endorsement (sample mean value of item set).

Table N2

Absolute Evaluation Criteria When Only Comparing the Item Sets with the Lowest and the Highest Endorsement (Negative Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{occ}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 3 items																
14	low	0.257	0.288	0.312	.246	.491	.971	.324	.378	.449	0	0	0	6	8.169	11
20	low	0.249	0.283	0.302	.551	.643	.732	.312	.361	.420	0	0	0	5	7.576	11
30	low	0.258	0.277	0.295	.726	.947	.997	.273	.335	.381	0	0	0	3	5.917	9
40	low	0.255	0.271	0.285	.776	.848	.99	.267	.313	.358	0	0	0	4	6.25	9
50	low	0.255	0.268	0.288	.833	.864	.893	.278	.316	.352	0	0	0	5	5.814	9
60	low	0.254	0.264	0.276	.871	.906	.994	.273	.300	.335	0	0	0	4	5.051	6
14	high	0.255	0.276	0.296	.592	.651	.725	.125	.175	.239	0	0	0	0	0.800	3
20	high	0.244	0.257	0.274	.659	.714	.779	.091	.144	.210	0	0	0	0	0.667	2
30	high	0.229	0.240	0.251	.755	.815	.986	.080	.107	.142	0	0	0	0	0.250	2
40	high	0.214	0.225	0.237	.796	.819	.849	.051	.083	.125	0	0	0	0	0.200	1
50	high	0.209	0.217	0.227	.829	.845	.866	.051	.071	.097	0	0	0	0	0.033	1
60	high	0.198	0.210	0.215	.839	.861	.874	.040	.062	.080	0	0	0	0	0.017	1
Item sets with 6 items																
14	low	0.150	0.178	0.206	.475	.771	.988	.136	.192	.273	0	0	0	0	0.117	2
20	low	0.150	0.174	0.202	.560	.716	.804	.102	.151	.199	0	0	0	0	0.467	2
30	low	0.146	0.163	0.186	.739	.799	.990	.074	.110	.153	0	0	0	0	0.183	1
40	low	0.140	0.157	0.172	.791	.836	.866	.051	.085	.125	0	0	0	0	0.133	1
50	low	0.145	0.156	0.166	.851	.869	.889	.045	.070	.108	0	0	0	0	0	0
60	low	0.141	0.152	0.159	.865	.884	.898	.040	.060	.085	0	0	0	0	0.017	1
14	high	0.190	0.204	0.224	.610	.769	.988	.028	.076	.136	0	0	0	0	0.068	1
20	high	0.172	0.190	0.207	.677	.738	.791	.023	.054	.085	0	0	0	0	0.250	2
30	high	0.162	0.177	0.192	.765	.804	.990	.017	.035	.057	0	0	0	0	0.085	1
40	high	0.155	0.168	0.178	.804	.836	.865	.017	.030	.062	0	0	0	0	0	0
50	high	0.152	0.162	0.169	.837	.860	.879	.017	.024	.034	0	0	0	0	0	0
60	high	0.150	0.158	0.163	.862	.876	.884	.017	.022	.034	0	0	0	0	0	0

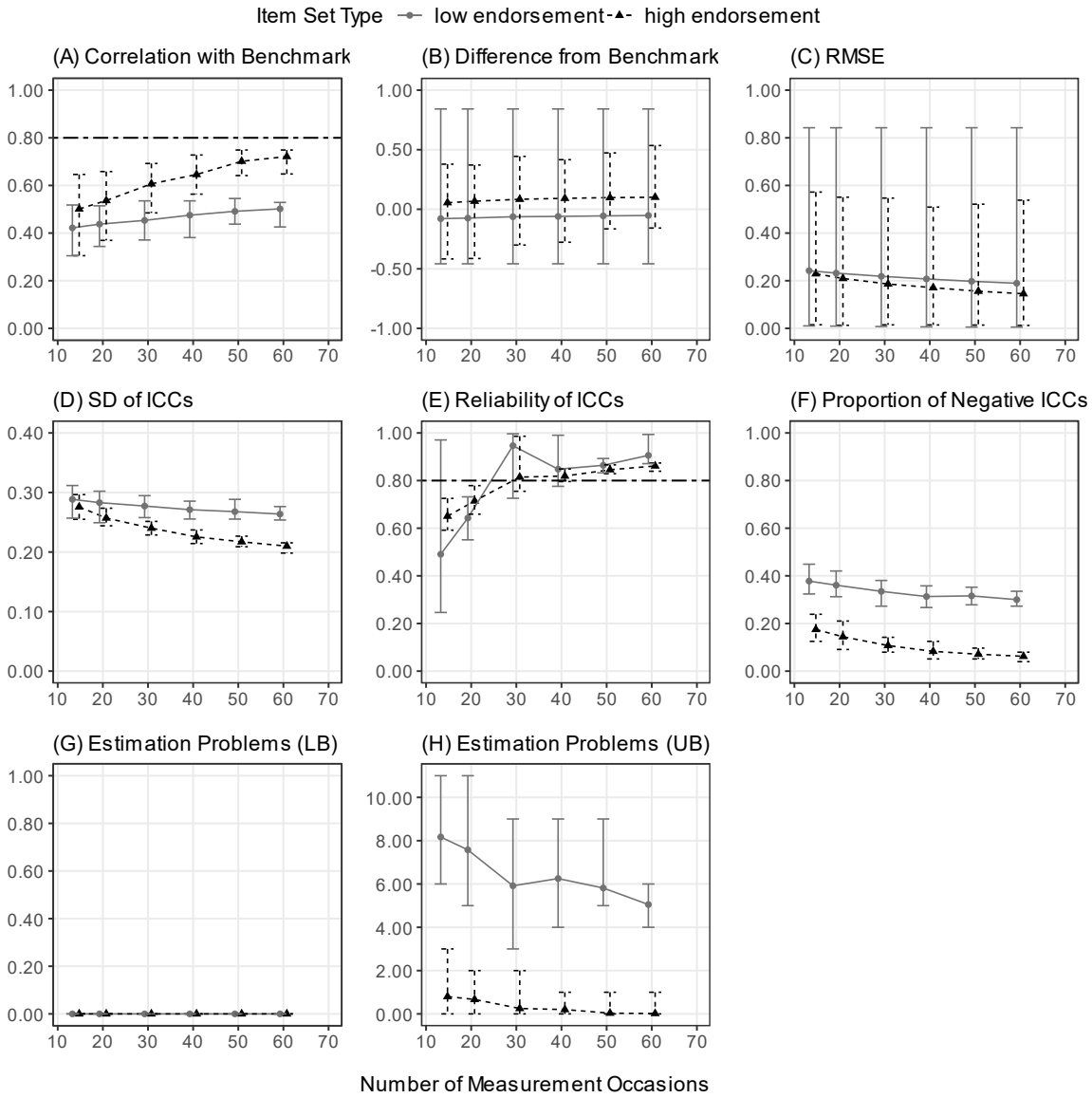
Table N2 (continued)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{occ}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 8 items																
14	low	0.146	0.172	0.203	.548	.748	.988	.028	.074	.136	0	0	0	0	0.049	1
20	low	0.139	0.165	0.189	.650	.740	.819	.017	.053	.097	0	0	0	0	0.245	2
30	low	0.133	0.155	0.175	.740	.817	.99	.006	.035	.074	0	0	0	0	0.013	1
40	low	0.131	0.149	0.168	.805	.844	.886	.000	.024	.062	0	0	0	0	0.034	2
50	low	0.132	0.145	0.157	.840	.867	.893	.000	.018	.040	0	0	0	0	0	0
60	low	0.128	0.141	0.152	.858	.883	.905	.000	.014	.034	0	0	0	0	0	0
14	high	0.152	0.178	0.204	.571	.742	.988	.023	.066	.136	0	0	0	0	0.045	1
20	high	0.145	0.170	0.195	.680	.750	.821	.011	.047	.091	0	0	0	0	0.214	2
30	high	0.134	0.160	0.180	.746	.826	.995	.006	.032	.074	0	0	0	0	0.020	1
40	high	0.138	0.154	0.170	.813	.851	.890	.000	.023	.051	0	0	0	0	0.029	1
50	high	0.136	0.149	0.162	.848	.872	.898	.006	.019	.051	0	0	0	0	0	0
60	high	0.135	0.146	0.154	.869	.888	.900	.006	.016	.028	0	0	0	0	0	0

Note. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 59 to 60 replications (3-item and 6-item sets) or 555 to 556 replications (8-item sets), respectively, per occasion-number condition. The mean values were *M* = 1.18 for the low endorsement 3-item set, *M* = 1.73 for the high endorsement 3-item set, *M* = 1.36 for the low endorsement 6-item set, *M* = 1.63 for the high endorsement 6-item set, *M* = 1.44 for the low endorsement 8-item set, and *M* = 1.52 for the high endorsement 8-item set. Since the 9-item conditions included the full item set, we could not compare low and high endorsement item sets for this number of items. The reliability calculation required Fisher's *Z*-transformed ICCs. RMSE = root mean square error of condition in reference to the benchmark. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. End. = endorsement (sample mean value of item set). Min = minimum. *M* = mean. Max = maximum. End. = endorsement (sample mean value of item set).

Figure N1

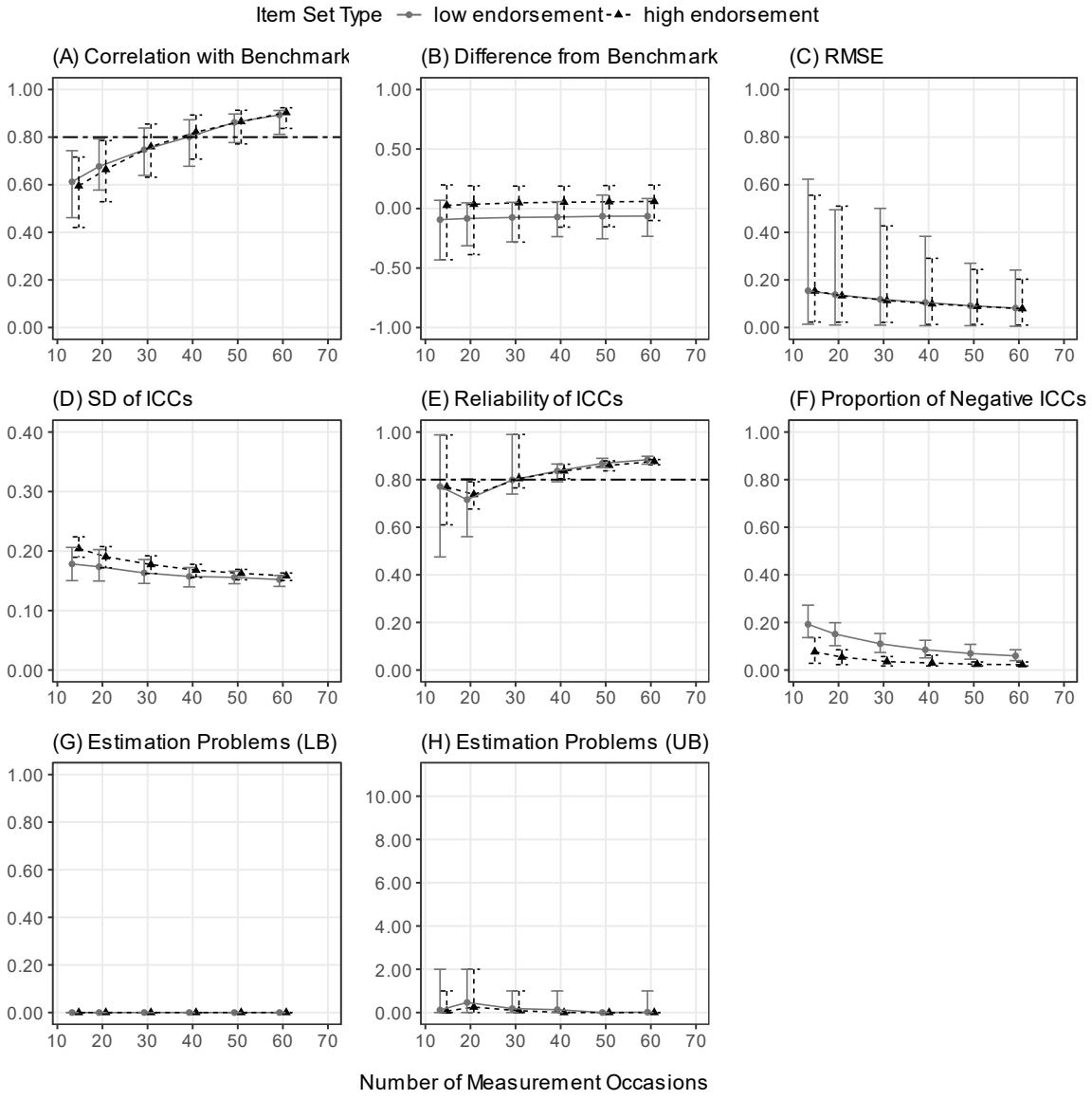
Results of the Evaluation Criteria for raw Intraclass Correlations (ICCs) Separate for the Item Sets with the Highest and the Lowest Endorsement in the Sample (3 Items, Negative Emotion Differentiation)



Note. Panels (A) to (H) depict six evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (H) are absolute criteria. The panels present results for 3 items for negative emotion differentiation (NED) in dependence of the item endorsement. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 59 to 60 replications per occasion-number condition. The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across replications. The plots display mean values for low and high endorsement item sets. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs. LB = lower bound. UB = upper bound.

Figure N2

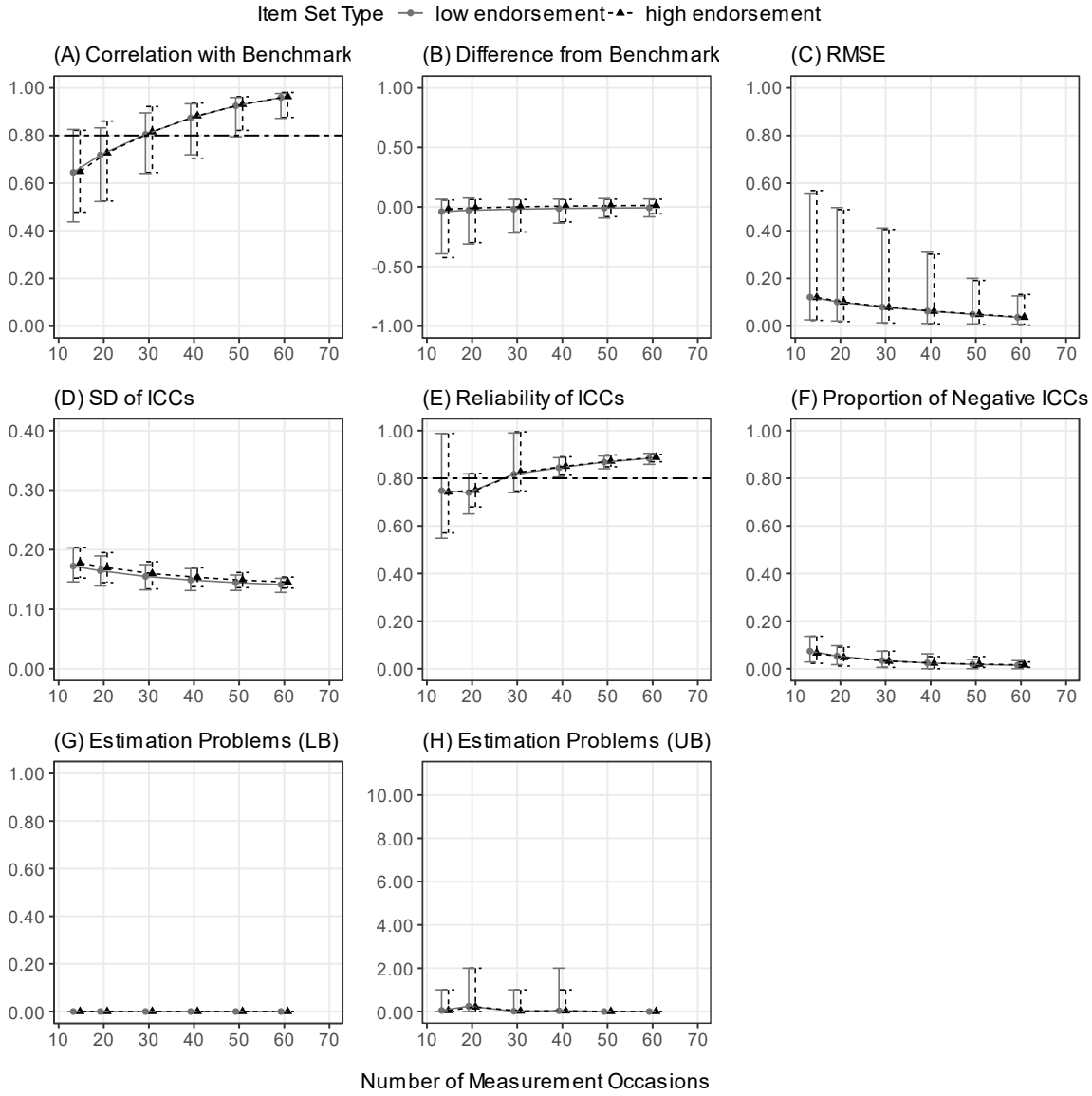
Results of the Evaluation Criteria for raw Intraclass Correlations (ICCs) Separate for the Item Sets with the Highest and the Lowest Endorsement in the Sample (6 Items, Negative Emotion Differentiation)



Note. Panels (A) to (H) depict six evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (H) are absolute criteria. The panels present results for 6 items for negative emotion differentiation (NED) in dependence of the item endorsement. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 59 to 60 replications per occasion-number condition. The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across replications. The plots display mean values for low and high endorsement item sets. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs. LB = lower bound. UB = upper bound.

Figure N3

Results of the Evaluation Criteria for raw Intraclass Correlations (ICCs) Separate for the Item Sets with the Highest and the Lowest Endorsement in the Sample (8 Items, Negative Emotion Differentiation)



Note. Panels (A) to (H) depict six evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (H) are absolute criteria. The panels present results for 8 items for negative emotion differentiation (NED) in dependence of the item endorsement. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 555 to 556 replications per occasion-number condition. The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across replications. The plots display mean values for low and high endorsement item sets. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs. LB = lower bound. UB = upper bound.

Table N3

Relative Evaluation Criteria When Only Comparing the Item Sets with the Lowest and the Highest Endorsement (Positive Emotion Differentiation in the Replication Data Set)

Condition		Correlation			Difference			RMSE		
N_{occ}	End.	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 3 items										
14	low	.384	.531	.658	-0.279	-0.030	0.468	0.089	0.175	0.503
20	low	.443	.594	.689	-0.273	-0.024	0.488	0.075	0.149	0.506
30	low	.572	.658	.729	-0.280	-0.019	0.511	0.061	0.123	0.516
40	low	.636	.699	.756	-0.275	-0.018	0.515	0.050	0.107	0.518
50	low	.679	.726	.768	-0.276	-0.017	0.519	0.037	0.096	0.520
60	low	.717	.746	.777	-0.275	-0.016	0.519	0.026	0.086	0.519
14	high	.418	.610	.708	-0.281	-0.007	0.278	0.117	0.177	0.387
20	high	.585	.678	.765	-0.292	-0.002	0.284	0.088	0.148	0.374
30	high	.650	.745	.807	-0.294	0.001	0.291	0.067	0.121	0.332
40	high	.729	.786	.830	-0.309	0.003	0.291	0.052	0.104	0.331
50	high	.781	.814	.850	-0.310	0.004	0.290	0.038	0.092	0.321
60	high	.796	.833	.856	-0.311	0.004	0.294	0.027	0.081	0.314

Table N3 (continued)

Condition		Correlation			Difference			RMSE		
N_{occ}	End.	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 4 items										
14	low	.472	.592	.703	-0.103	0.059	0.504	0.074	0.158	0.543
20	low	.521	.655	.750	-0.097	0.065	0.545	0.058	0.138	0.557
30	low	.649	.722	.790	-0.091	0.070	0.561	0.041	0.117	0.563
40	low	.698	.762	.811	-0.084	0.072	0.564	0.033	0.106	0.565
50	low	.748	.791	.828	-0.084	0.074	0.564	0.024	0.097	0.564
60	low	.787	.811	.834	-0.084	0.075	0.566	0.019	0.090	0.566
14	high	.537	.684	.779	-0.122	0.049	0.167	0.096	0.153	0.298
20	high	.622	.758	.832	-0.086	0.054	0.171	0.073	0.129	0.220
30	high	.750	.833	.884	-0.095	0.058	0.175	0.059	0.107	0.200
40	high	.825	.878	.912	-0.097	0.059	0.183	0.043	0.093	0.196
50	high	.871	.908	.933	-0.094	0.061	0.183	0.031	0.083	0.190
60	high	.910	.929	.945	-0.093	0.061	0.185	0.020	0.075	0.188

Note. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 500 replications (3-item sets) or 1,000 replications (4-item sets), respectively, per occasion-number condition. The mean values were $M = 2.13$ for the low endorsement 3-item set, $M = 3.03$ for the high endorsement 3-item set, $M = 2.29$ for the low endorsement 4-item set, and $M = 2.85$ for the high endorsement 4-item set. Since the 5-item conditions included the full item set, we could not compare low and high endorsement item sets for this number of items. The reliability calculation required Fisher's *Z*-transformed ICCs. Correlation = correlation of the ICCs between condition and benchmark. Difference = person-level difference in ICCs between condition and benchmark. RMSE = root mean square error of condition in reference to the benchmark. N_{occ} = number of measurement occasions. End. = endorsement (sample mean value of item set). Min = minimum. *M* = mean. Max = maximum. End. = endorsement (sample mean value of item set).

Table N4

Absolute Evaluation Criteria When Only Comparing the Item Sets with the Lowest and the Highest Endorsement (Positive Emotion Differentiation in the Replication Data Set)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{occ}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 3 items																
14	low	0.182	0.208	0.233	.101	.322	.502	.045	.094	.159	0	0	0	0	0.070	2
20	low	0.166	0.189	0.208	.232	.442	.596	.017	.065	.131	0	0	0	0	0.024	1
30	low	0.153	0.172	0.187	.410	.558	.646	.006	.037	.074	0	0	0	0	0.002	1
40	low	0.151	0.163	0.178	.558	.636	.701	.000	.024	.062	0	0	0	0	0	0
50	low	0.146	0.158	0.169	.626	.688	.738	.000	.016	.040	0	0	0	0	0	0
60	low	0.147	0.153	0.159	.699	.727	.760	.000	.011	.028	0	0	0	0	0	0
14	high	0.201	0.228	0.263	.253	.443	.618	.045	.103	.165	0	0	0	0	0	0
20	high	0.181	0.208	0.228	.366	.540	.650	.017	.075	.142	0	0	0	0	0	0
30	high	0.175	0.190	0.208	.562	.637	.706	.017	.051	.097	0	0	0	0	0	0
40	high	0.167	0.182	0.196	.640	.703	.754	.011	.041	.074	0	0	0	0	0	0
50	high	0.165	0.176	0.187	.710	.748	.779	.006	.035	.068	0	0	0	0	0	0
60	high	0.165	0.172	0.180	.762	.780	.803	.011	.030	.051	0	0	0	0	0	0

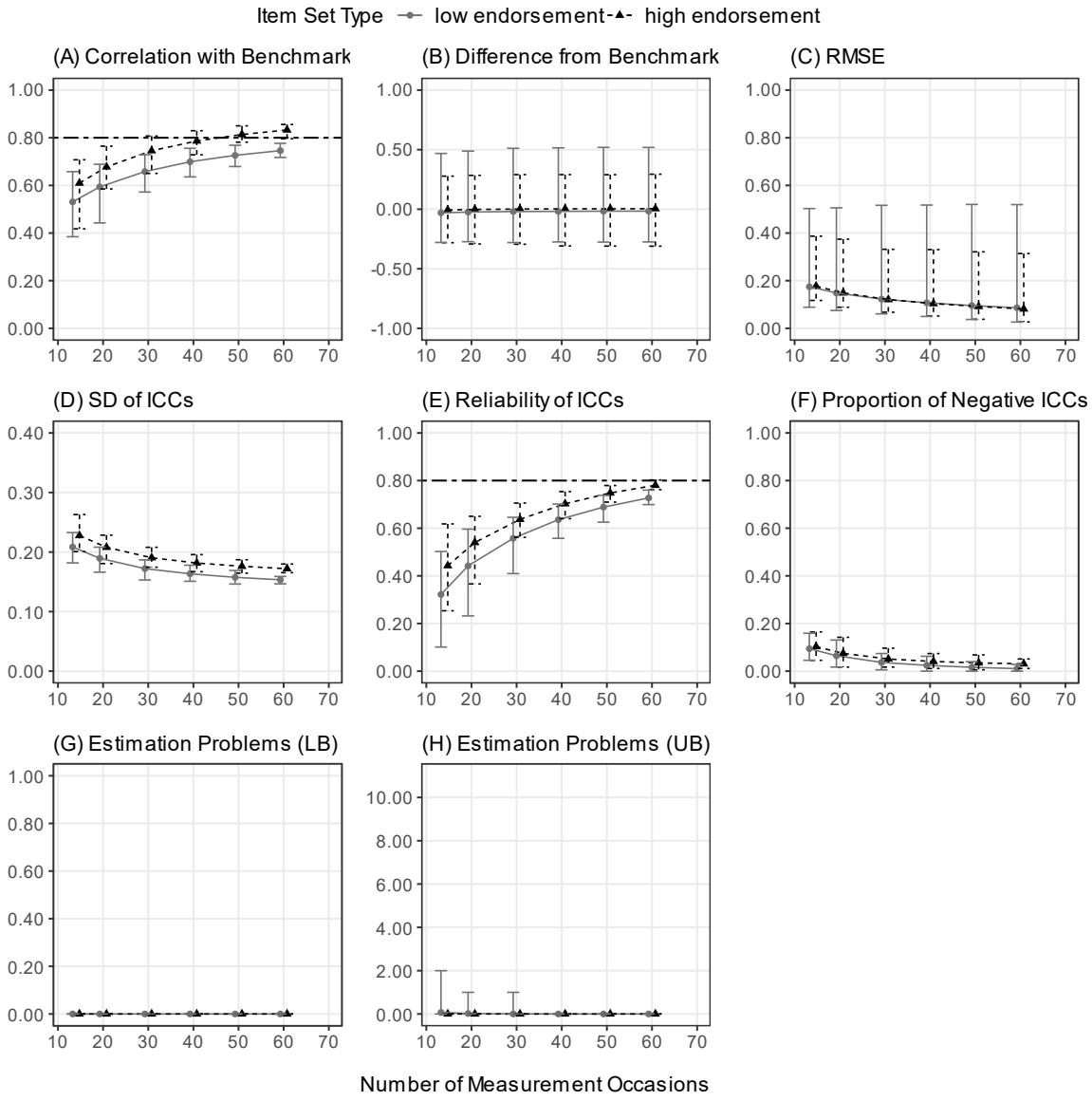
Table N4 (continued)

Condition		*SD*			Reliability			% Neg ICC			Estimation Problems					
											LB			UB		
N_{occ}	N_{occ}	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max	Min	*M*	Max
Item sets with 4 items																
14	low	0.166	0.191	0.220	.213	.440	.610	.000	.029	.080	0	0	0	0	0.070	2
20	low	0.152	0.174	0.195	.359	.543	.678	.000	.014	.051	0	0	0	0	0.015	1
30	low	0.142	0.158	0.176	.531	.639	.988	.000	.005	.023	0	0	0	0	0	0
40	low	0.138	0.150	0.163	.640	.702	.774	.000	.002	.011	0	0	0	0	0	0
50	low	0.134	0.145	0.154	.690	.745	.785	.000	.000	.011	0	0	0	0	0	0
60	low	0.136	0.142	0.148	.758	.778	.801	.000	.000	.006	0	0	0	0	0	0
14	high	0.174	0.201	0.227	.292	.494	.653	.006	.042	.091	0	0	0	0	0	0
20	high	0.165	0.183	0.205	.458	.581	.676	.000	.026	.062	0	0	0	0	0	0
30	high	0.150	0.169	0.184	.570	.674	.740	.000	.017	.051	0	0	0	0	0	0
40	high	0.148	0.161	0.173	.681	.732	.777	.000	.012	.034	0	0	0	0	0	0
50	high	0.144	0.155	0.164	.731	.772	.800	.000	.010	.023	0	0	0	0	0	0
60	high	0.147	0.152	0.160	.786	.803	.823	.000	.010	.017	0	0	0	0	0	0

Note. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 500 replications (3-item sets) or 1,000 replications (4-item sets), respectively, per occasion-number condition. The mean values were *M* = 2.13 for the low endorsement 3-item set, *M* = 3.03 for the high endorsement 3-item set, *M* = 2.29 for the low endorsement 4-item set, and *M* = 2.85 for the high endorsement 4-item set. Since the 5-item conditions included the full item set, we could not compare low and high endorsement item sets for this number of items. The reliability calculation required Fisher's *Z*-transformed ICCs. *SD* = standard deviation of ICCs across participants. % Neg ICC = proportion of participants with an ICC below zero. N_{occ} = number of measurement occasions. End. = endorsement (sample mean value of item set). Min = minimum. *M* = mean. Max = maximum. End. = endorsement (sample mean value of item set).

Figure N4

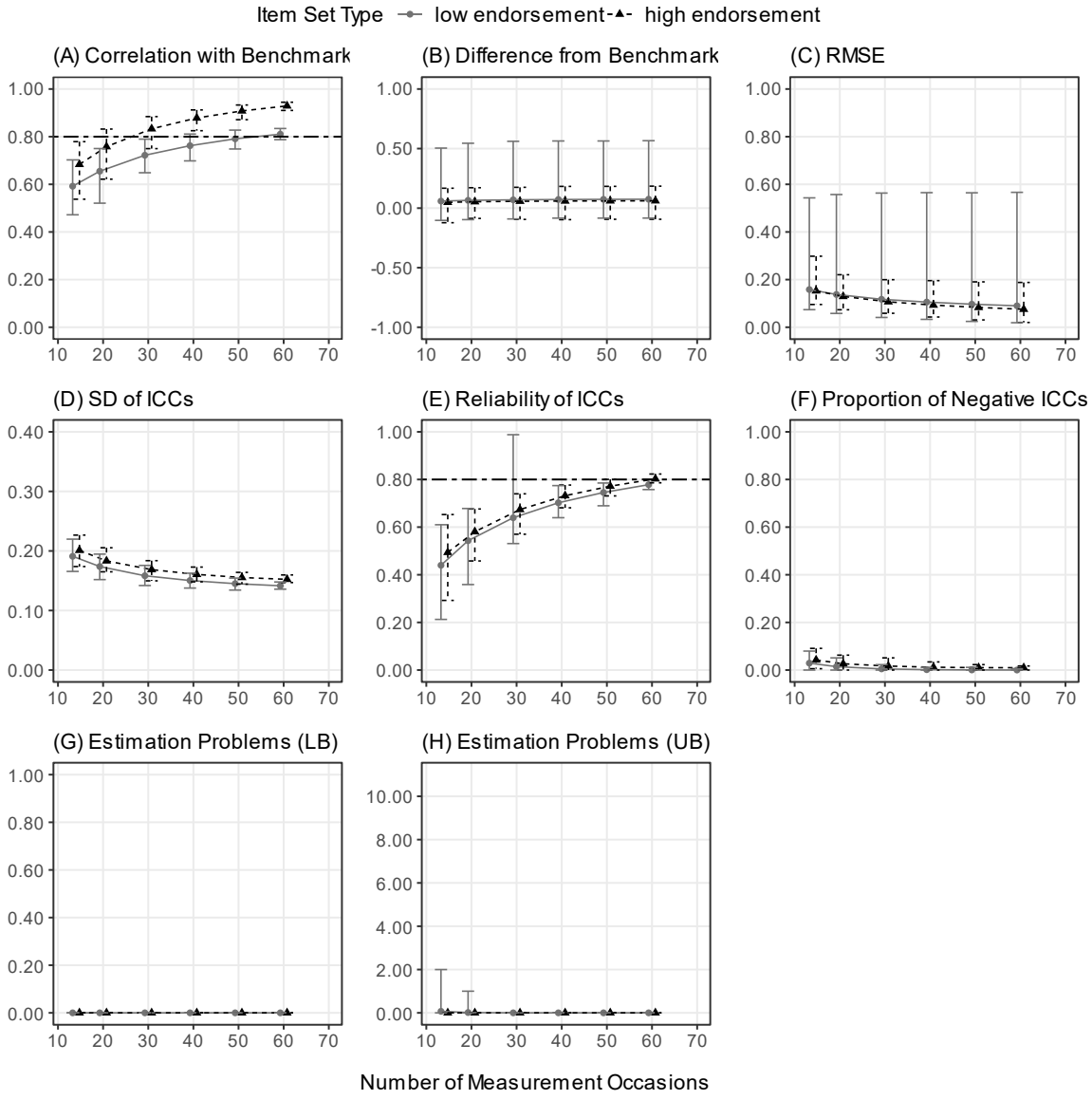
Results of the Evaluation Criteria for raw Intraclass Correlations (ICCs) Separate for the Item Sets with the Highest and the Lowest Endorsement in the Sample (3 Items, Positive Emotion Differentiation)



Note. Panels (A) to (H) depict six evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (H) are absolute criteria. The panels present results for 3 items for positive emotion differentiation (PED) in dependence of the item endorsement. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 500 replications per occasion-number condition. The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across replications. The plots display mean values for low and high endorsement item sets. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs. LB = lower bound. UB = upper bound.

Figure N5

Results of the Evaluation Criteria for raw Intraclass Correlations (ICCs) Separate for the Item Sets with the Highest and the Lowest Endorsement in the Sample (4 Items, Positive Emotion Differentiation)



Note. Panels (A) to (H) depict six evaluation criteria: (A) to (C) are relative to the benchmark condition; (D) to (H) are absolute criteria. The panels present results for 4 items for positive emotion differentiation (PED) in dependence of the item endorsement. Since all possible item sets were implemented and balanced across 5,000 replications, the item sets compared here were used in only 1,000 replications per occasion-number condition. The difference from the benchmark and the root mean squared error (RMSE) were calculated per participant (across replications) and then averaged; the other evaluation criteria were averaged across replications. The plots display mean values for low and high endorsement item sets. Error bars in (B) and (C) indicate person-level minima and maxima; those in the remaining panels reflect minimum and maximum values across replications. Horizontal dashed lines mark .80 in (A) and (E). Reliability uses Fisher's *Z*-transformed ICCs; all other criteria use raw ICCs. LB = lower bound. UB = upper bound.

References

Bicaker, E., Lane, S. P., Sadikaj, G., & Racine, S. E. (2022). The roles of negative emotion intensity, negative emotion differentiation, and self-compassion in loss of control eating. *The International Journal of Eating Disorders*, *55*(7), 966–976. https://doi.org/10.1002/eat.23723

Boden, M. T., Thompson, R. J., Dizén, M., Berenbaum, H., & Baker, J. P. (2013). Are emotional clarity and emotion differentiation related? *Cognition & Emotion*, *27*(6), 961–978. https://doi.org/10.1080/02699931.2012.751899

Bresin, K. (2014). Five indices of emotion regulation in participants with a history of nonsuicidal self-injury: A daily diary study. *Behavior Therapy*, *45*(1), 56–66. https://doi.org/10.1016/j.beth.2013.09.005

Brown, B. A., Goodman, F. R., Disabato, D. J., Kashdan, T. B., Armeli, S., & Tennen, H. (2021). Does negative emotion differentiation influence how people choose to regulate their distress after stressful events? A four-year daily diary study. *Emotion*, *21*(5), 1000–1012. https://doi.org/10.1037/emo0000969

Decker, M. L., Turk, C. L., Hess, B., & Murray, C. E. (2008). Emotion regulation among individuals classified with and without generalized anxiety disorder. *Journal of Anxiety Disorders*, *22*(3), 485–494. https://doi.org/10.1016/j.janxdis.2007.04.002

Demiralp, E., Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Feldman Barrett, L., Ellsworth, P. C., Demiralp, M., Hernandez-Garcia, L., Deldin, P. J., Gotlib, I. H., & Jonides, J. (2012). Feeling blue or turquoise? Emotional differentiation in major depressive disorder. *Psychological Science*, *23*(11), 1410–1416. https://doi.org/10.1177/0956797612444903

Dixon-Gordon, K. L., Chapman, A. L., Weiss, N. H., & Rosenthal, M. Z. (2014). A preliminary examination of the role of emotion differentiation in the relationship between borderline personality and urges for maladaptive behaviors. *Journal of*

Psychopathology and Behavioral Assessment, *36*(4), 616–625. https://doi.org/10.1007/s10862-014-9423-4

D'Mello, S. K., & Gruber, J. (2021). Emotional regularity: Associations with personality, psychological health, and occupational outcomes. *Cognition & Emotion*, *35*(8), 1460–1478. https://doi.org/10.1080/02699931.2021.1968797

Dong, B., & Xu, G. (2022). An empirical study on the evaluation of emotional complexity in daily life. *Frontiers in Psychology*, *13*, 839133. https://doi.org/10.3389/fpsyg.2022.839133

Eckland, N. S., Sperry, S. H., Castro, A. A., & Berenbaum, H. (2022). Intensity, frequency, and differentiation of discrete emotion categories in daily life and their associations with depression, worry, and rumination. *Emotion, 22*(2), 305–317. https://doi.org/10.1037/emo0001038

Emery, N. N., Simons, J. S., Clarke, C. J., & Gaher, R. M. (2014). Emotion differentiation and alcohol-related problems: The mediating role of urgency. *Addictive Behaviors, 39*(10), 1459–1463. https://doi.org/10.1016/j.addbeh.2014.05.004

Emery, N. N., Walters, K. J., Njeim, L., Barr, M., Gelman, D., & Eddie, D. (2022). Emotion differentiation in early recovery from alcohol use disorder: Associations with in-the-moment affect and 3-month drinking outcomes. *Alcoholism, Clinical and Experimental Research*, *46*(7), 1294–1305. https://doi.org/10.1111/acer.14854

Erbas, Y., Ceulemans, E., Koval, P., & Kuppens, P. (2015). The role of valence focus and appraisal overlap in emotion differentiation. *Emotion*, *15*(3), 373–382. https://doi.org/10.1037/emo0000039

Erbas, Y., Ceulemans, E., Lee Pe, M., Koval, P., & Kuppens, P. (2014). Negative emotion differentiation: Its personality and well-being correlates and a comparison of different assessment methods. *Cognition & Emotion*, *28*(7), 1196–1213. https://doi.org/10.1080/02699931.2013.875890

Erbas, Y., Sels, L., Ceulemans, E., & Kuppens, P. (2016). Feeling me, feeling you: The relation between emotion differentiation and empathic accuracy. *Social Psychological and Personality Science, 7*(3), 240–247. https://doi.org/10.1177/1948550616633504

Feldman Barrett, L., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition & Emotion*, *15*(6), 713–724. https://doi.org/10.1080/02699930143000239

Grühn, D., Lumley, M. A., Diehl, M., & Labouvie-Vief, G. (2013). Time-based indicators of emotional complexity: Interrelations and correlates. *Emotion*, *13*(2), 226–237. https://doi.org/10.1037/a0030363

Hill, C. L. M., & Updegraff, J. A. (2012). Mindfulness and its relationship to emotional regulation. *Emotion*, *12*(1), 81–90. https://doi.org/10.1037/a0026355

Hoemann, K., Khan, Z., Kamona, N., Dy, J., Barrett, L. F., & Quigley, K. S. (2021). Investigating the relationship between emotional granularity and cardiorespiratory physiological activity in daily life. *Psychophysiology*, *58*(6), e13818. https://doi.org/10.1111/psyp.13818

Hoemann, K., Lee, Y., Kuppens, P., Gendron, M., & Boyd, R. L. (2023). Emotional granularity is associated with daily experiential diversity. *Affective Science*, *4*(2), 291–306. https://doi.org/10.1007/s42761-023-00185-2

Hohensee, N., Joormann, J., & Gadassi-Polack, R. (2024). Examining the association among adolescents' emotional clarity, emotion differentiation, and the regulation of negative and positive affect using a daily diary approach. *Emotion.* Advance online publication. https://doi.org/10.1037/emo0001424

Kalokerinos, E. K., Erbas, Y., Ceulemans, E., & Kuppens, P. (2019). Differentiate to regulate: Low negative emotion differentiation is associated with ineffective use but not

selection of emotion-regulation strategies. *Psychological Science*, *30*(6), 863–879. https://doi.org/10.1177/0956797619838763

Kashdan, T. B., DeWall, C. N., Masten, C. L., Pond, R. S., Powell, C., Combs, D., Schurtz, D. R., & Farmer, A. S. (2014). Who is most vulnerable to social rejection? The toxic combination of low self-esteem and lack of negative emotion differentiation on neural responses to rejection. *PloS One*, *9*(3), e90651. https://doi.org/10.1371/journal.pone.0090651

Kashdan, T. B., & Farmer, A. S. (2014). Differentiating emotions across contexts: Comparing adults with and without social anxiety disorder using random, social interaction, and daily experience sampling. *Emotion*, *14*(3), 629–638. https://doi.org/10.1037/a0035796

Kashdan, T. B., Ferssizidis, P., Collins, R. L., & Muraven, M. (2010). Emotion differentiation as resilience against excessive alcohol use: An ecological momentary assessment in underage social drinkers. *Psychological Science*, *21*(9), 1341–1347. https://doi.org/10.1177/0956797610379863

Kimhy, D., Vakhrusheva, J., Khan, S., Chang, R. W., Hansen, M. C., Ballon, J. S., Malaspina, D., & Gross, J. J. (2014). Emotional granularity and social functioning in individuals with schizophrenia: An experience sampling study. *Journal of Psychiatric Research*, *53*, 141–148. https://doi.org/10.1016/j.jpsychires.2014.01.020

Knapp, K. S., Bradizza, C. M., Zhao, J., Linn, B. K., Wilding, G. E., LaBarre, C., & Stasiewicz, P. R. (2024). Emotion differentiation among individuals in a randomized clinical trial for alcohol use disorder: Within- and between-person associations with affect, craving, and alcohol use in daily life. *Behaviour Research and Therapy*, *173*, 104474. https://doi.org/10.1016/j.brat.2024.104474

Krafft, J., Ong, C. W., Cruz, R. A., Twohig, M. P., & Levin, M. E. (2020). An ecological momentary assessment study investigating the function of hoarding. *Behavior Therapy*, *51*(5), 715–727. https://doi.org/10.1016/j.beth.2019.10.006

Lazarus, G., & Fisher, A. J. (2021). Negative emotion differentiation predicts psychotherapy outcome: Preliminary findings. *Frontiers in Psychology*, *12*, 689407. https://doi.org/10.3389/fpsyg.2021.689407

Lennarz, H. K., Lichtwarck-Aschoff, A., Timmerman, M. E., & Granic, I. (2018). Emotion differentiation and its relation with emotional well-being in adolescents. *Cognition & Emotion*, *32*(3), 651–657. https://doi.org/10.1080/02699931.2017.1338177

Liao, L., Zhang, K., Zhou, Y., & Liu, J. (2024). The association between negative emotion differentiation and emotion regulation flexibility in daily life. *Cognition & Emotion*, 1–13. https://doi.org/10.1080/02699931.2024.2381079

Lischetzke, T., Schemer, L., Glombiewski, J. A., In-Albon, T., Karbach, J., & Könen, T. (2021). Negative emotion differentiation attenuates the within-person indirect effect of daily stress on nightly sleep quality through calmness. *Frontiers in Psychology*, *12*, 684117. https://doi.org/10.3389/fpsyg.2021.684117

Liu, D. Y., Gilbert, K. E., & Thompson, R. J. (2020). Emotion differentiation moderates the effects of rumination on depression: A longitudinal study. *Emotion*, *20*(7), 1234–1243. https://doi.org/10.1037/emo0000627

MacCann, C., Erbas, Y., Dejonckheere, E., Minbashian, A., Kuppens, P., & Fayn, K. (2020). Emotional intelligence relates to emotions, emotion dynamics, and emotion complexity. *European Journal of Psychological Assessment*, *36*(3), 460–470. https://doi.org/10.1027/1015-5759/a000588

Mikhail, M. E., Keel, P. K., Burt, S. A., Neale, M., Boker, S., & Klump, K. L. (2020). Low emotion differentiation: An affective correlate of binge eating? *The International Journal of Eating Disorders*, *53*(3), 412–421. https://doi.org/10.1002/eat.23207

Oh, V. Y. S., & Tong, E. M. W. (2020). Negative emotion differentiation and long-term physical health - The moderating role of neuroticism. *Health Psychology*, *39*(2), 127–136. https://doi.org/10.1037/hea0000809

O'Toole, M. S., Jensen, M. B., Fentz, H. N., Zachariae, R., & Hougaard, E. (2014). Emotion differentiation and emotion regulation in high and low socially anxious individuals: An experience-sampling study. *Cognitive Therapy and Research*, *38*(4), 428–438. https://doi.org/10.1007/s10608-014-9611-2

Ottenstein, C., & Lischetzke, T. (2020). Development of a novel method of emotion differentiation that uses open-ended descriptions of momentary affective states. *Assessment*, *27*(8), 1928–1945. https://doi.org/10.1177/1073191119839138

Pond, R. S., Kashdan, T. B., DeWall, C. N., Savostyanova, A., Lambert, N. M., & Fincham, F. D. (2012). Emotion differentiation moderates aggressive tendencies in angry people: A daily diary analysis. *Emotion*, *12*(2), 326–337. https://doi.org/10.1037/a0025762

Pugach, C. P., Starr, L. R., Silvia, P. J., & Wisco, B. E. (2023). Negative emotion differentiation in trauma-exposed community members: Associations with posttraumatic stress disorder symptoms in daily life. *Journal of Psychopathology and Clinical Science*, *132*(8), 1007–1018. https://doi.org/10.1037/abn0000851

Ryvkina, E., Kroencke, L., Geukes, K., Scharbert, J., & Back, M. D. (2023). Understanding psychological responses to the COVID-19 pandemic through ESM data: The EMOTIONS project. *Journal of Open Psychology Data*, *11*, 6. https://doi.org/10.5334/jopd.83

Schreuder, M. J., Wichers, M., Hartman, C. A., Menne-Lothmann, C., Decoster, J., van Winkel, R., Delespaul, P., Hert, M. de, Derom, C., Thiery, E., Rutten, B. P. F., Jacobs, N., & van Os, J. (2022). Lower emotional complexity as a prospective

predictor of psychopathology in adolescents from the general population. *Emotion*, *22*(5), 836–843. https://doi.org/10.1037/emo0000778

Seah, T. H. S., Aurora, P., & Coifman, K. G. (2020). Emotion differentiation as a protective factor against the behavioral consequences of rumination: A conceptual replication and extension in the context of social anxiety. *Behavior Therapy*, *51*(1), 135–148. https://doi.org/10.1016/j.beth.2019.05.011

Seah, T. H. S., & Coifman, K. G. (2024). Effects of scaffolding emotion language use on emotion differentiation and psychological health: An experience-sampling study. *Cognition & Emotion*, 1–18. https://doi.org/10.1080/02699931.2024.2382334

Seah, T. H. S., Sidney, P. G., Taber, J. M., Thompson, C. A., & Coifman, K. G. (2023). Emotional complexity under high stress: Do protective associations for risk behaviors persist even during a pandemic? *Emotion*, *23*(3), 879–885. https://doi.org/10.1037/emo0001133

Selby, E. A., Wonderlich, S. A., Crosby, R. D., Engel, S. G., Panza, E., Mitchell, J. E., Crow, S. J., Peterson, C. B., & Le Grange, D. (2014). Nothing tastes as good as thin feels: Low positive emotion differentiation and weight loss activities in anorexia nervosa. *Clinical Psychological Science*, *2*(4), 514–531. https://doi.org/10.1177/2167702613512794

Sels, L., Erbas, Y., O'Brien, S. T., Verhofstadt, L., Clark, M. S., & Kalokerinos, E. K. (2024). The double-edged sword of social sharing: Social sharing predicts increased emotion differentiation when rumination is low but decreased emotion differentiation when rumination is high. *Psychological Science*, *35*(10), 1079–1093. https://doi.org/10.1177/09567976241266513

Sheets, E. S., Bujarski, S., Leventhal, A. M., & Ray, L. A. (2015). Emotion differentiation and intensity during acute tobacco abstinence: A comparison of heavy and light

smokers. *Addictive Behaviors, 47*, 70–73. https://doi.org/10.1016/j.addbeh.2015.03.024

Siepe, B. S., Bartoš, F., Morris, T. P., Boulesteix, A.-L., Heck, D. W., & Pawel, S. (2024). Simulation studies for methodological research in psychology: A standardized template for planning, preregistration, and reporting. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000695

Starr, L. R., Hershenberg, R., Li, Y. I., & Shaw, Z. A. (2017). When feelings lack precision: Low positive and negative emotion differentiation and depressive symptoms in daily life. *Clinical Psychological Science*, *5*(4), 613–631. https://doi.org/10.1177/2167702617694657

Starr, L. R., Hershenberg, R., Shaw, Z. A., Li, Y. I., & Santee, A. C. (2020). The perils of murky emotions: Emotion differentiation moderates the prospective relationship between naturalistic stress exposure and adolescent depression. *Emotion*, *20*(6), 927–938. https://doi.org/10.1037/emo0000630

Thompson, R. J., Liu, D. Y., Sudit, E., & Boden, M. (2021). Emotion differentiation in current and remitted major depressive disorder. *Frontiers in Psychology*, *12*, 685851. https://doi.org/10.3389/fpsyg.2021.685851

Tomko, R. L., Lane, S. P., Pronove, L. M., Treloar, H. R., Brown, W. C., Solhan, M. B., Wood, P. K., & Trull, T. J. (2015). Undifferentiated negative affect and impulsivity in borderline personality and depressive disorders: A momentary perspective. *Journal of Abnormal Psychology*, *124*(3), 740–753. https://doi.org/10.1037/abn0000064

Tong, E. M. W., & Keng, S.-L. (2017). The relationship between mindfulness and negative emotion differentiation: A test of multiple mediation pathways. *Mindfulness*, *8*(4), 933–942. https://doi.org/10.1007/s12671-016-0669-7

Tugade, M. M., Fredrickson, B. L., & Feldman Barrett, L. (2004). Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on

coping and health. *Journal of Personality*, *72*(6), 1161–1190. https://doi.org/10.1111/j.1467-6494.2004.00294.x

van der Gucht, K., Dejonckheere, E., Erbas, Y., Takano, K., Vandemoortele, M., Maex, E., Raes, F., & Kuppens, P. (2019). An experience sampling study examining the potential impact of a mindfulness-based intervention on emotion differentiation. *Emotion*, *19*(1), 123–131. https://doi.org/10.1037/emo0000406

Vandercammen, L., Hofmans, J., & Theuns, P. (2014). Relating specific emotions to intrinsic motivation: On the moderating role of positive and negative emotion differentiation. *PloS One*, *9*(12), e115396. https://doi.org/10.1371/journal.pone.0115396

Walukevich-Dienst, K., Piccirillo, M. L., Calhoun, B. H., Bedard-Gilligan, M., Larimer, M. E., Patrick, M. E., & Lee, C. M. (2023). Daily-level relationships between negative affect, negative emotion differentiation, and cannabis behaviors among a high-risk sample of young adults. *Journal of Affective Disorders*, *335*, 392–400. https://doi.org/10.1016/j.jad.2023.05.056

Widdershoven, R. L. A., Wichers, M., Kuppens, P., Hartmann, J. A., Menne-Lothmann, C., Simons, C. J. P., & Bastiaansen, J. A. (2019). Effect of self-monitoring through experience sampling on emotion differentiation in depression. *Journal of Affective Disorders*, *244*, 71–77. https://doi.org/10.1016/j.jad.2018.10.092

Williams-Kerver, G. A., & Crowther, J. H. (2020). Emotion differentiation and disordered eating behaviors: The role of appearance schemas. *Eating Behaviors*, *37*, 101369. https://doi.org/10.1016/j.eatbeh.2020.101369

Willroth, E. C., Flett, J. A. M., & Mauss, I. B. (2020). Depressive symptoms and deficits in stress-reactive negative, positive, and within-emotion-category differentiation: A daily diary study. *Journal of Personality*, *88*(2), 174–184. https://doi.org/10.1111/jopy.12475

Yang, Y. (2022). The association of daily stressors and daily emotions to daily optimism. *Journal of Individual Differences*, *43*(4), 216–227. https://doi.org/10.1027/1614-0001/a000375

Yue, P., An, Z., & Lin, R. (2024). How do "words poorly expressed emotion" affect mental health? The mediating role of affect labelling effect. *Cognition & Emotion*, 1–16. https://doi.org/10.1080/02699931.2024.2362377

Zaki, L. F., Coifman, K. G., Rafaeli, E., Berenson, K. R., & Downey, G. (2013). Emotion differentiation as a protective factor against nonsuicidal self-injury in borderline personality disorder. *Behavior Therapy*, *44*(3), 529–540. https://doi.org/10.1016/j.beth.2013.04.008