Evaluating the factor structure of the Dutch

Individual Differences in Language Skills (IDLaS-NL) test battery

Florian Hintz[1,2,3], Marjolijn Dijkhuis[2], Vera van 't Hoff[2], Milou Huijsmans[2], Rogier A. Kievit[4,5], James M. McQueen[4,2], and Antje S. Meyer[2,4]

[1]Philipps University Marburg, [2]Max Planck Institute for Psycholinguistics, [3]Center for Mind, Brain and Behavior, Philipps University Marburg & Justus Liebig University Giessen, [4]Donders Institute for Brain, Cognition, and Behaviour, [5]Radboud University Medical Centre

Supplementary Information B:

Descriptive statistics and factor score distributions

Table 1: Descriptive statistics based on participants who completed the battery in the laboratory ('in-lab participants', n = 169).

Domain	Task	N[a]	Incorrect (%)	Trimmed (%)[bc]	Mean (SD)	Range	Skewness[d]	Kurtosis[d]	IC
Linguistic knowledge	Peabody Picture Vocabulary Test (PPVT)	169	–	–	175.88 (9.89)	136 – 198	-0.60	0.94	0.89[fg]
	Antonym production (Antonyms)	169	–	–	0.80 (0.07)	0.52 – 0.96	-0.31	0.62	0.44[h]
	Idiom recognition (Idioms)	169	–	–	0.77 (0.11)	0.40 – 1.0	-0.50	0.28	0.19[f]
	Spelling test (Spelling)	169	–	–	0.58 (0.17)	0.10 – 0.93	-0.28	-0.37	0.77[f]
	Dutch Author Recognition Test (DART)	166	–	–	0.26 (0.14)	0.02 – 0.80	1.05	1.24	0.93[f]
	Prescriptive grammar test (Grammar)	169	–	–	0.69 (0.10)	0.425 – 0.925	0.00	-0.29	0.61[f]
General cognitive skills	Auditory simple RT test (A.SRT)	168	–	0.77%	Inv. log: -2.31 (0.06) Raw: 212 (34)	Inv. log: -2.52 – -2.16 Raw: 146 – 356	-0.31[i]	0.51[i]	0.90[ih]
	Auditory choice RT test (A.CRT)	165	3.52%	1.65%	Inv. log: -2.56 (0.08) Raw: 384 (86)	Inv. log: -2.92 – -2.42 Raw: 265 – 915	-1.10[i]	1.84[i]	0.96[ih]
	Letter comparison (LetterComp)	169	6.16%	0.71%	Inv. log: -2.98 (0.07) Raw: 1061 (186)	Inv. log: -3.19 – -2.81 Raw: 687 – 1680	-0.08[i]	-0.07[i]	0.97[ih]
	Visual simple RT test (V.SRT)	169	–	0.27%	Inv. log: -2.36 (0.05) Raw: 236 (29)	Inv. log: -2.54 – -2.25 Raw: 178 – 360	-0.67[i]	0.51[i]	0.87[ih]
	Visual choice RT test (V.CRT)	168	3.56%	0.06%	Inv. log: -2.60 (0.06) Raw: 408 (59)	Inv. log: -2.77 – -2.47 Raw: 297 – 624	-0.40[i]	0.23[i]	0.93[ih]
	Digit span: forward (DigitSpanForward) Digit span: backward (DigitSpanBackward)	169 168	– –	– –	8.51 (1.76) 7.79 (2.05)	4 – 14 1 – 12	0.29 0.03	-0.66 -0.40	0.71[fj] 0.71[fj]
	Corsi block test: forward (CorsiForward) Corsi block test: backward (CorsiBackward)	166 168	– –	– –	8.27 (1.93) 7.60 (2.03)	4 – 13 2 – 14	0.09 0.23	-0.05 0.41	0.60[fj] 0.69[fj]
	Raven's Advanced Progressive Matrices Test (Raven)	166	–	–	0.60 (0.15)	0.22 – 0.94	-0.13	-0.33	0.82[fj]
Linguistic processing skills	Picture naming (PicNam/PN)	159	7.23%	–	Inv. log: -2.96 (0.06) Raw: 953 (140)	Inv. log: -3.18 – -2.81 Raw: 661 – 1580	-0.46[i]	0.69[i]	0.95[ih]
	Rapid Automatized Naming (RAN)	166	0.08%[k]	–	1.60 (0.23)	1.00 – 2.29	0.16	-0.12	0.95[h]
	Verbal fluency: categories (VerbalFluencyCat/VFC) Verbal fluency: letters (VerbalFluencyPhon/VFP)	166 168	1.78%[k] 0.59%[k]	– –	24.95 (4.63) 15.33 (4.08)	14.5 – 38.5 6 – 24	0.30 -0.10	-0.33 -0.56	0.63[e] 0.60[e]
	Maximal speech rate (MaxSpeechRate/MSR)	164	7.62%[l]	–	Inv. log: -3.67 (0.09) Raw: 4815 (1113)	Inv. log: -3.94 – -3.45 Raw: 2815 – 8657	-0.52[i]	0.06[i]	0.89[ie]
	Phrase generation (PhraseGen/PhG)	166	2.77%	–	Inv. log: -3.16 (0.06) Raw: 1516 (203)	Inv. log: -3.34 – -3.02 Raw: 1071 – 2256	-0.06[i]	-0.08[i]	0.98[ih]
	Sentence generation structured (sSentenceGen/sSG)	153	6.99%	–	Inv. log: -3.42 (0.06) Raw: 2716 (399)	Inv. log: -3.59 – -3.30 Raw: 1983 – 4032	-0.42[i]	0.14[i]	0.98[ih]
	Sentence generation unstructured (fSentenceGen/fSG)	155	6.47%	–	Inv. log: -2.82 (0.06) Raw: 702 (103)	Inv. log: -2.98 – -2.63 Raw: 441 – 995	0.08[i]	-0.10[i]	0.96[ih]
	Non-word monitoring in noise in lists (NonwordMonitoring/NwM)	165	–	–	0.52 (0.13)	0.00 – 0.81	-1.04	1.87	0.44[f]
	Rhyme judgment (RhymeJudge/RJ)	166	2.28%	0.51%	Inv. log: -2.90 (0.09) Raw: 828 (186)	Inv. log: -3.14 – -2.72 Raw: 535 – 1440	-0.52[i]	-0.04[i]	0.95[ih]

	Auditory lexical decision (A.LDT)	167	2.58%	0.19%	Inv. log: -2.94 (0.05) Raw: 897 (124)	Inv. log: -3.10 – -2.84 Raw: 693 – 1338	-0.64[i]	0.31[i]	0.97[ih]
	Semantic categorization (SemCat/SC)	169	2.20%	0.15%	Inv. log: -2.93 (0.05) Raw: 877 (121)	Inv. log: -3.17 – -2.81 Raw: 655 – 1494	-0.72[i]	1.61[i]	0.95[ih]
	Word monitoring in noise in sentences (SentenceMonitoring/SM)	167	–	–	0.54 (0.13)	0.0 – 0.8	-0.59	0.70	0.43[f]
	Gender cue activation during sentence comprehension (MorphPred/MP)	165	1.55%	0.18%	-691 (618)	-1795 – 638	-0.26	-0.93	0.94[h]
	Verb semantics activation during sentence comprehension (SemanticPred/SP)	164	0.09%	0.34%	-885 (567)	-1829 – 580	-0.80	-0.09	0.93[h]
	Self-paced reading (SPR)	155	–	5.95%[o]	Inv. log: -2.63 (0.10) Raw: 441 (99)	Inv. log: -2.88 – -2.23 Raw: 170 – 771	0.44[i]	1.20[i]	0.99[ih]

Note:
[a] See Methods section in the main manuscript for missing and excluded values.
[b] See Table 2 in the main manuscript for the lower and upper boundaries for the reaction time trimming procedure.
[c] If applicable, based on correct responses.
[d] Calculated based on aggregated performance indicators.
[e] The Spearman-Brown split-half reliability is calculated using the spearman_brown() function of the splithalfr R package. If applicable, participants with only 1 valid trial or run are removed from calculation.
[f] Cronbach's Alpha is calculated using the cronbach.alpha() function of the ltm R package.
[g] Unseen easier trials (trialnumbers below 145) are coded as correct; unseen harder trials (trialnumbers above 145) are coded as incorrect.
[h] ICC3k is calculated using the ICC() function of the psych R package.
[i] Calculated based on inverse log-transformed values.
[j] Unseen trials are coded as incorrect.
[k] Incorrect trials here represent trials where participants did not take the trial seriously or did not understand the instructions.
[l] For participants who had one incorrect trial, descriptive statistics are calculated based on the speech rate of the one remaining correct trial.
[m] Active and other responses are coded as non-passive in order to establish a dichotomous response variable.
[n] Incorrect trials here represent trials with speech duration lower than 30 seconds.
[o] Trimmed trials here represent trials where at least one word in the sentence was trimmed.

Table 2: Descriptive statistics based on participants who completed the battery remotely via the internet ('remote participants', n = 579).

Domain	Task	N[a]	Incorrect (%)	Trimmed (%)[bc]	Mean (SD)	Range	Skewness[d]	Kurtosis[d]	IC
Linguistic knowledge	Peabody Picture Vocabulary Test (PPVT)	578	–	–	176.26 (10.68)	141 – 198	-0.74	0.33	0.90[fg]
	Antonym production (Antonyms)	579	–	–	0.78 (0.08)	0.36 – 0.96	-0.62	1.05	0.49[h]
	Idiom recognition (Idioms)	579	–	–	0.77 (0.13)	0.3 – 1.0	-0.40	-0.05	0.37[f]
	Spelling test (Spelling)	578	–	–	0.63 (0.16)	0.13 – 1.00	-0.28	-0.30	0.75[f]
	Dutch Author Recognition Test (DART)	578	–	–	0.28 (0.15)	0.00 – 0.81	0.89	0.61	0.94[f]
	Prescriptive grammar test (Grammar)	579	–	–	0.69 (0.11)	0.375 – 0.975	0.05	-0.34	0.67[f]
General cognitive skills	Auditory simple RT test (A.SRT)	579	–	0.76%	Inv. log: -2.51 (0.10) Raw: 336 (85)	Inv. log: -2.83 – -2.30 Raw: 202 – 674	-1.04[i]	0.65[i]	0.98[ih]
	Auditory choice RT test (A.CRT)	577	4.58%	0.23%	Inv. log: -2.68 (0.09) Raw: 499 (123)	Inv. log: -3.08 – -2.50 Raw: 317 – 1234	-0.94[i]	0.92[i]	0.98[ih]
	Letter comparison (LetterComp)	570	6.27%	0.77%	Inv. log: -3.01 (0.07) Raw: 1127 (190)	Inv. log: -3.27 – -2.83 Raw: 681 – 1941	0.02[i]	0.09[i]	0.97[ih]
	Visual simple RT test (V.SRT)	577	–	0.68%	Inv. log: -2.44 (0.05) Raw: 284 (38)	Inv. log: -2.64 – -2.31 Raw: 208 – 479	-0.61[i]	0.79[i]	0.93[ih]
	Visual choice RT test (V.CRT)	579	3.55%	0.13%	Inv. log: -2.65 (0.06) Raw: 460 (72)	Inv. log: -2.90 – -2.50 Raw: 327 – 858	-0.82[i]	1.49[i]	0.95[ih]
	Digit span: forward (DigitSpanForward) Digit span: backward (DigitSpanBackward)	577 576	– –	– –	8.82 (2.10) 7.28 (2.27)	4 – 14 2 – 12	0.25 0.03	-0.50 -0.79	0.71[fj] 0.75[fj]
	Corsi block test: forward (CorsiForward) Corsi block test: backward (CorsiBackward)	577 577	– –	– –	8.27 (1.93) 7.60 (2.03)	1 – 14 1 – 14	-0.16 -0.05	0.37 0.48	0.67[fj] 0.68[fj]
	Raven's Advanced Progressive Matrices Test (Raven)	579	–	–	0.56 (0.14)	0.06 – 0.92	-0.50	0.48	0.80[fj]
Linguistic processing skills	Picture naming (PicNam/PN)	548	6.20%	–	Inv. log: -3.00 (0.05) Raw: 1026 (125)	Inv. log: -3.14 – -2.81 Raw: 663 – 1407	0.02[i]	-0.16[i]	0.94[ih]
	Rapid Automatized Naming (RAN)	573	1.80%[k]	–	1.58 (0.24)	0.88 – 2.27	-0.03	-0.21	0.95[h]
	Verbal fluency: categories (VerbalFluencyCat/VFC) Verbal fluency: letters (VerbalFluencyPhon/VFP)	563 562	2.43%[k] 2.77%[k]	– –	25.24 (4.85) 15.76 (4.32)	10.0 – 39.5 5 – 32	0.10 0.22	0.10 0.14	0.67[e] 0.72[e]
	Maximal speech rate (MaxSpeechRate/MSR)	555	6.41%[l]	–	Inv. log: -3.66 (0.09) Raw: 4693 (1049)	Inv. log: -4.10 – -3.44 Raw: 2721 – 12672	-0.80[i]	1.55[i]	0.92[ie]
	Phrase generation (PhraseGen/PhG)	561	2.81%	–	Inv. log: -3.21 (0.06) Raw: 1684 (233)	Inv. log: -3.42 – -3.03 Raw: 1086 – 2823	-0.40[i]	0.67[i]	0.99[ih]
	Sentence generation structured (sSentenceGen/sSG)	507	8.01%	–	Inv. log: -3.44 (0.05) Raw: 2823 (353)	Inv. log: -3.63 – -3.30 Raw: 2004 – 4336	-0.48[i]	0.66[i]	0.97[ih]
	Sentence generation unstructured (fSentenceGen/fSG)	493	–	–	Inv. log: -2.83 (0.06) Raw: 709 (104)	Inv. log: -3.06 – -2.68 Raw: 484 – 1200	-0.41[i]	0.58[i]	0.96[ih]
	Non-word monitoring in noise in lists (NonwordMonitoring/NwM)	553	–	–	0.46 (0.16)	0.00 – 0.78	-0.63	0.20	0.59[f]
	Rhyme judgment (RhymeJudge/RJ)	566	3.42%	0.47%	Inv. log: -2.97 (0.08) Raw: 980 (187)	Inv. log: -3.24 – -2.80 Raw: 636 – 1807	-0.49[i]	0.24[i]	0.94[ih]

	Auditory lexical decision (A.LDT)	578	2.71%	0.21%	Inv. log: -3.00 (0.05) Raw: 1031 (128)	Inv. log: -3.23 – -2.90 Raw: 796 – 1731	-0.65[i]	0.69[i]	0.97[ih]
	Semantic categorization (SemCat/SC)	575	2.20%	0.27%	Inv. log: -3.00 (0.06) Raw: 1018 (153)	Inv. log: -3.27 – -2.87 Raw: 750 – 1908	-0.69[i]	0.81[i]	0.97[ih]
	Word monitoring in noise in sentences (SentenceMonitoring/SM)	568	–	–	0.45 (0.16)	0.0 – 0.8	-0.52	-0.10	0.58[f]
	Gender cue activation during sentence comprehension (MorphPred/MP)	556	1.33%	0.55%	-500 (644)	-1640 – 921	-0.32	-0.93	0.95[h]
	Verb semantics activation during sentence comprehension (SemanticPred/SP)	574	0.41%	0.52%	-579 (615)	-1757 – 962	-0.49	-0.68	0.94[h]
	Self-paced reading (SPR)	549	0.01%[o]	4.13%[p]	Inv. log: -2.55 (0.11) Raw: 374 (95)	Inv. log: -2.83 – -2.21 Raw: 164 – 680	0.13[i]	-0.07[i]	0.99[ih]

Note:
[a] See Methods section in the main manuscript for missing and excluded values.
[b] See Table 2 in the main manuscript for the lower and upper boundaries for the reaction time trimming procedure.
[c] If applicable, based on correct responses.
[d] Calculated based on aggregated performance indicators.
[e] The Spearman-Brown split-half reliability is calculated using the spearman_brown() function of the splithalfr R package. If applicable, participants with only 1 valid trial or run are removed from calculation.
[f] Cronbach's Alpha is calculated using the cronbach.alpha() function of the ltm R package.
[g] Unseen easier trials (trialnumbers below 145) are coded as correct; unseen harder trials (trialnumbers above 145) are coded as incorrect.
[h] ICC3k is calculated using the ICC() function of the psych R package.
[i] Calculated based on inverse log-transformed values.
[j] Unseen trials are coded as incorrect.
[k] Incorrect trials here represent trials where participants did not take the trial seriously or did not understand the instructions.
[l] For participants who had one incorrect trial, descriptive statistics are calculated based on the speech rate of the one remaining correct trial.
[m] Active and other responses are coded as non-passive in order to establish a dichotomous response variable.
[n] Incorrect trials here represent trials with speech duration lower than 30 seconds.
[o] Incorrect trials here represent unreliable trials due to technical error.
[p] Trimmed trials here represent trials where at least one word in the sentence was trimmed.

Figure 1

Heatmap featuring Pearson's correlations between all performance indicators. Colder (i.e., blue) colours reflect increasingly more positive correlations and warmer (i.e., red) colours reflect increasingly more negative correlations. The correlations included in the seven highlighted squares belong to the same construct (i.e., Linguistic knowledge, Processing speed, Working memory, and Linguistic processing skills including, separately, Word production, Sentence production, Word comprehension and Sentence comprehension). *Panel A*: In-lab participants (n = 169). *Panel B*: Remote participants (n = 579). *Panel C:* Difference in correlation coefficients (in-lab minus remote). White squares in Panel C reflect no difference in correlation strength between in-lab and remote participant samples.

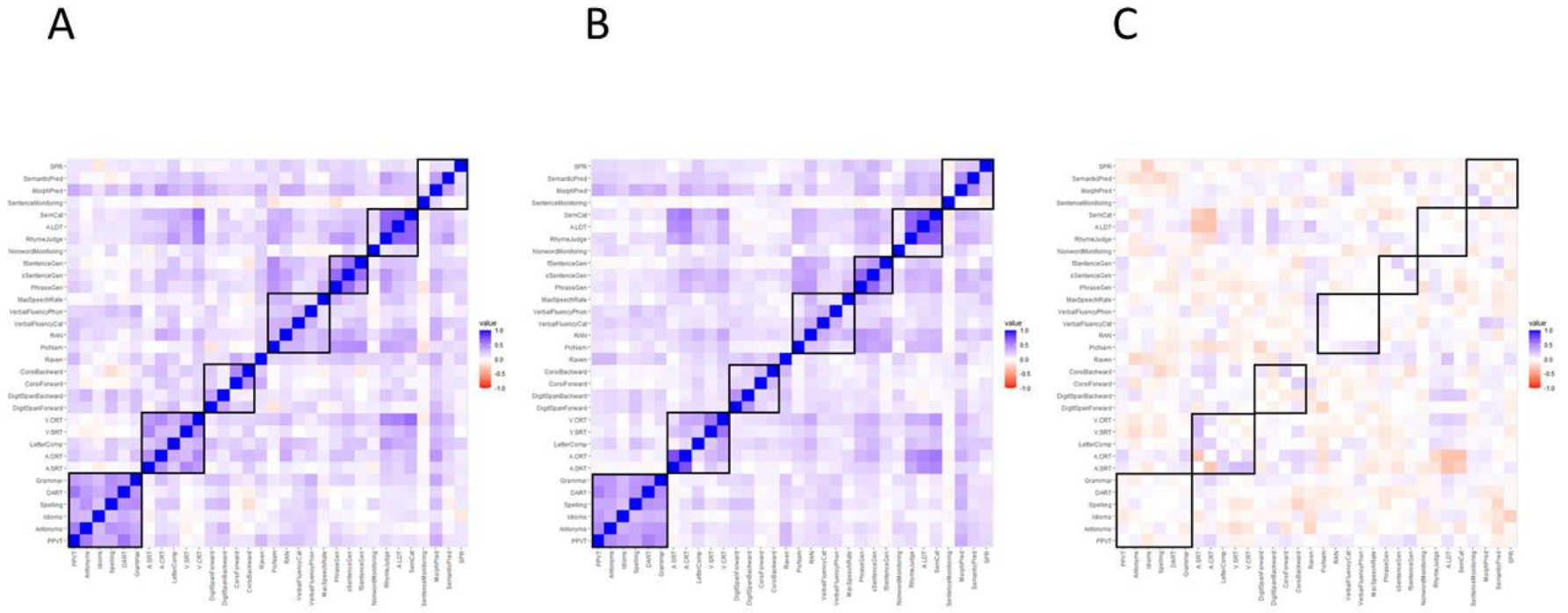

Figure 2

Score distributions for and correlations between *Linguistic knowledge* performance indicators for in-lab (red) and remote (green) experimental settings and for the combined data set.

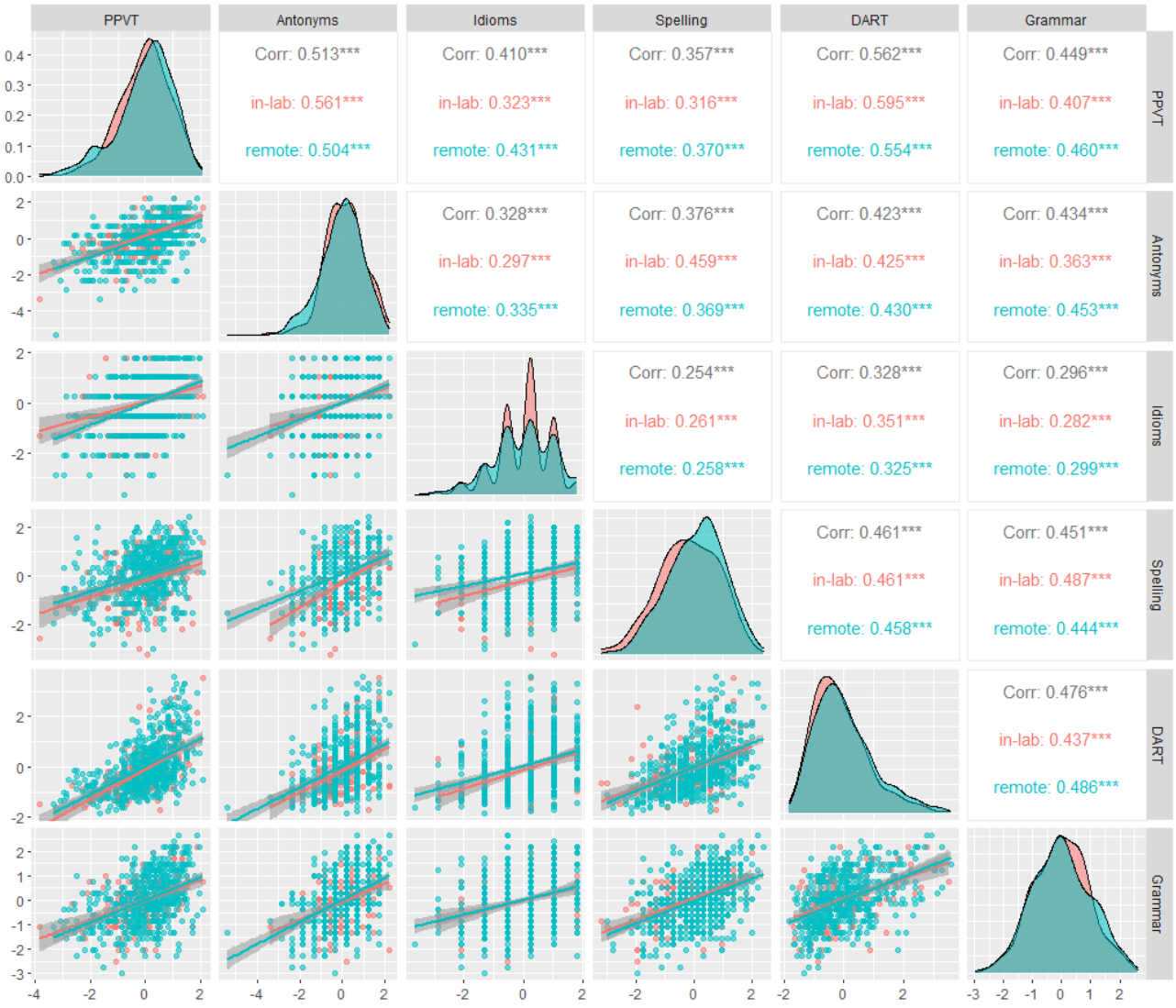

Figure 3
Distribution and frequency plots for predicted *Linguistic knowledge* factor scores across in-lab (red) and remote (green) experimental settings.

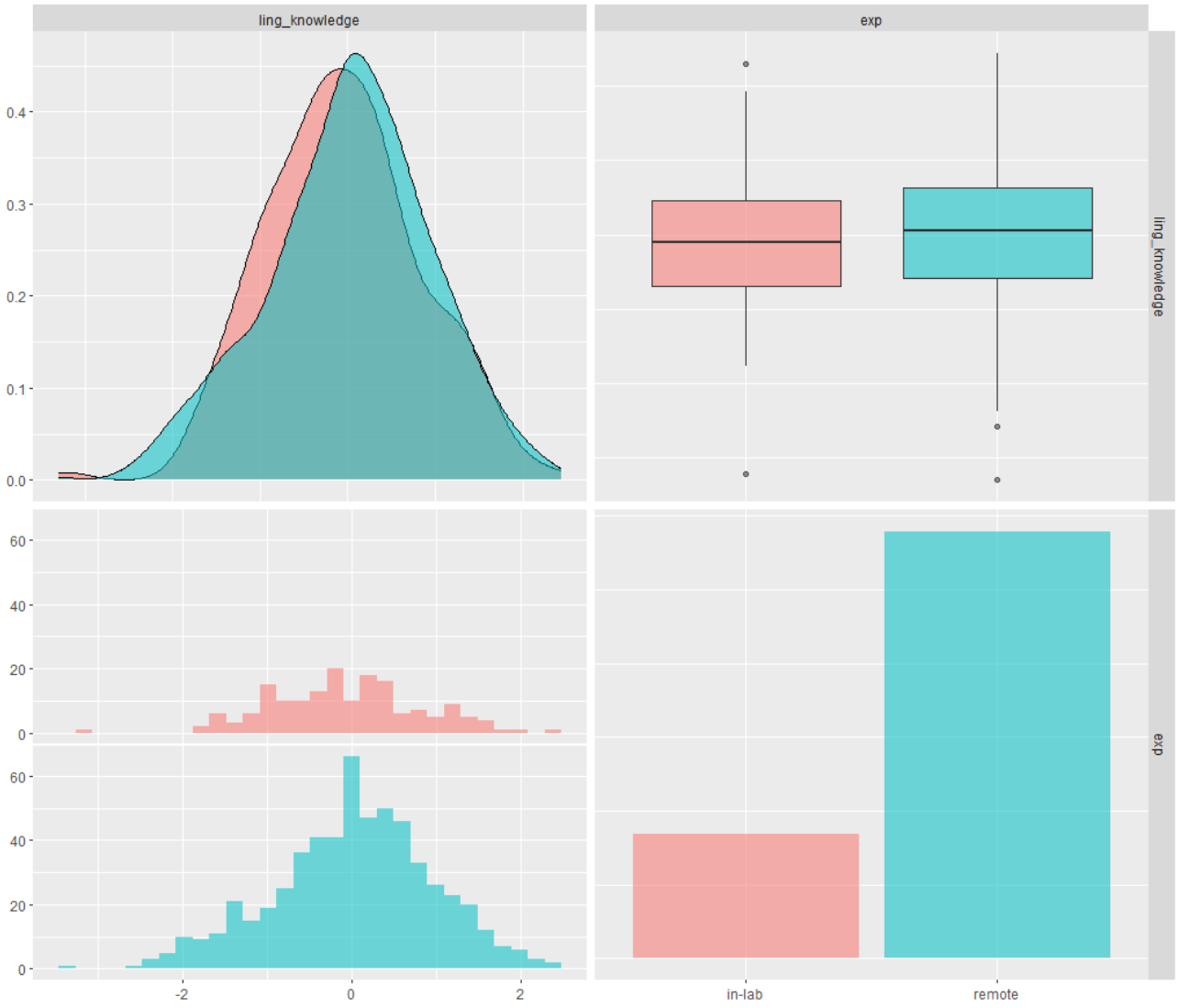

Figure 4

Score distributions for and correlations between *Processing speed* performance indicators for in-lab (red) and remote (green) experimental settings and for the combined data set.

Figure 5

Distribution and frequency plots for predicted *Visual and Auditory processing speed* factor scores across in-lab (red) and remote (green) experimental settings.

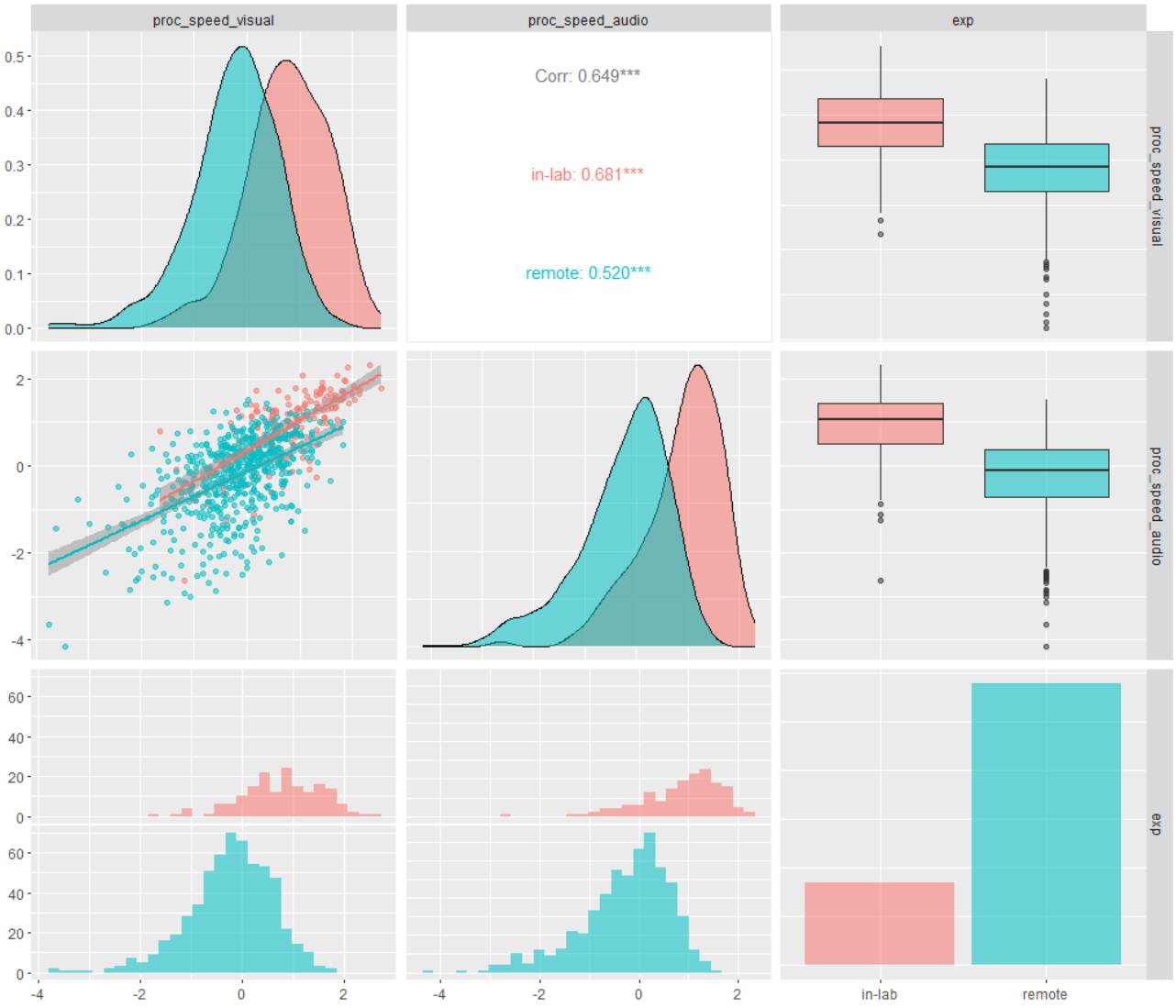

Figure 6

Score distributions for and correlations between *Working memory* performance indicators for in-lab (red) and remote (green) experimental settings and for the combined data set.

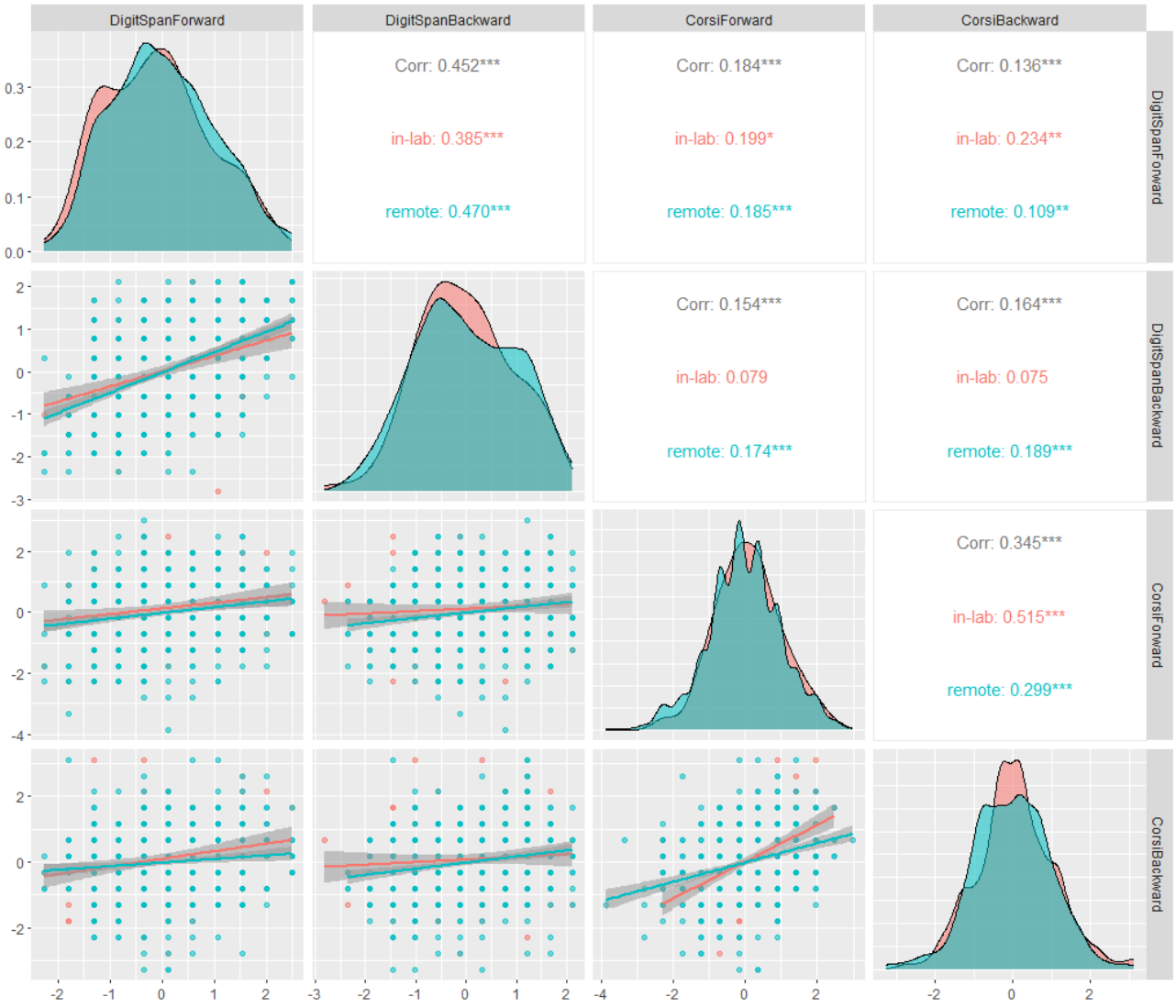

Figure 7

Distribution and frequency plots for predicted *Working memory* factor scores across in-lab (red) and remote (green) experimental settings.

Figure 8

Score distributions for and correlations between *Word production* (*Linguistic processing skills*) performance indicators for in-lab (red) and remote (green) experimental settings and for the combined data set.

Figure 9

Score distributions for and correlations between *Sentence production* (*Linguistic processing skills*) performance indicators for in-lab (red) and remote (green) experimental settings and for the combined data set.

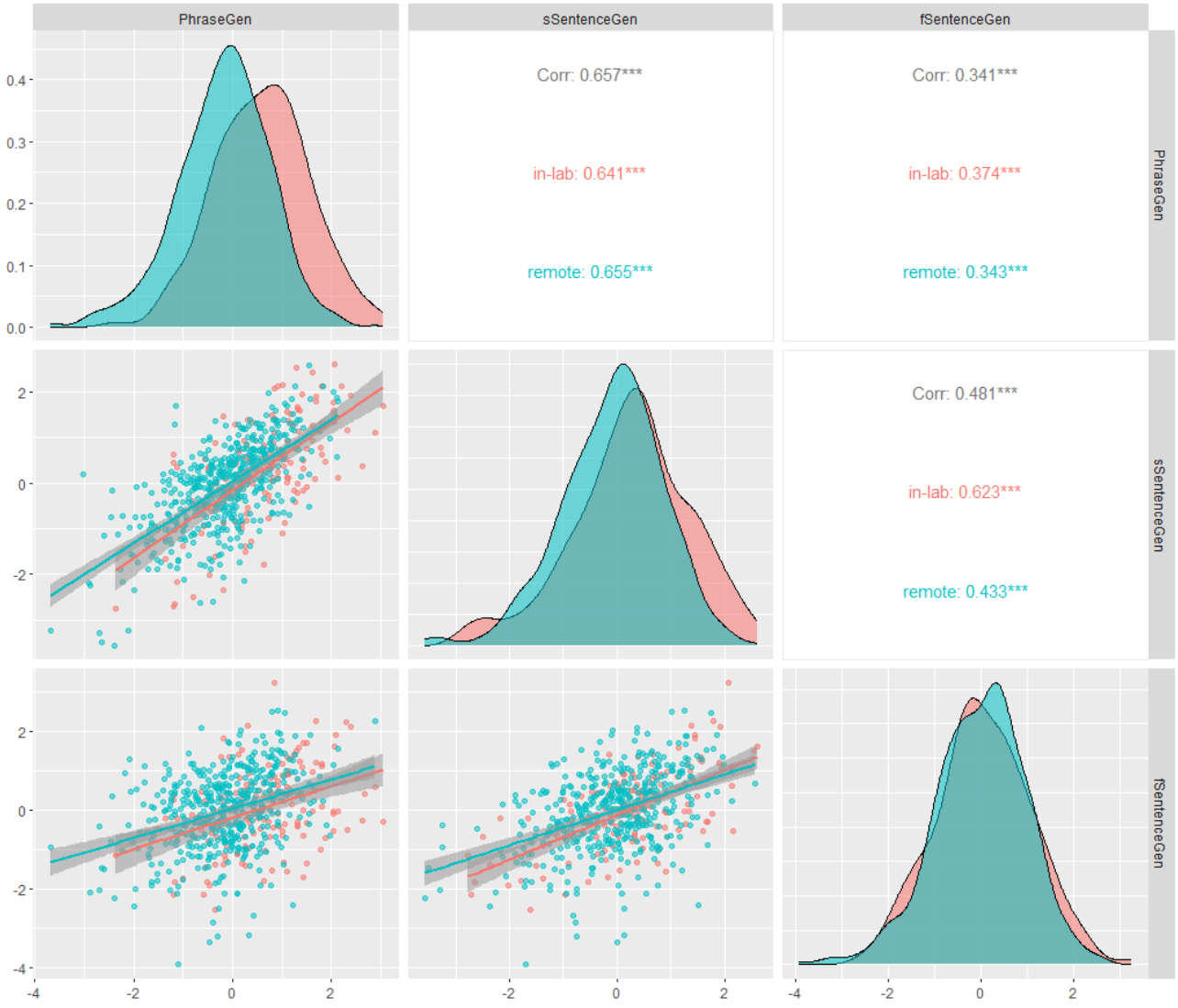

Figure 10

Score distributions for and correlations between *Word comprehension* (*Linguistic processing skills*) performance indicators for in-lab (red) and remote (green) experimental settings and for the combined data set.

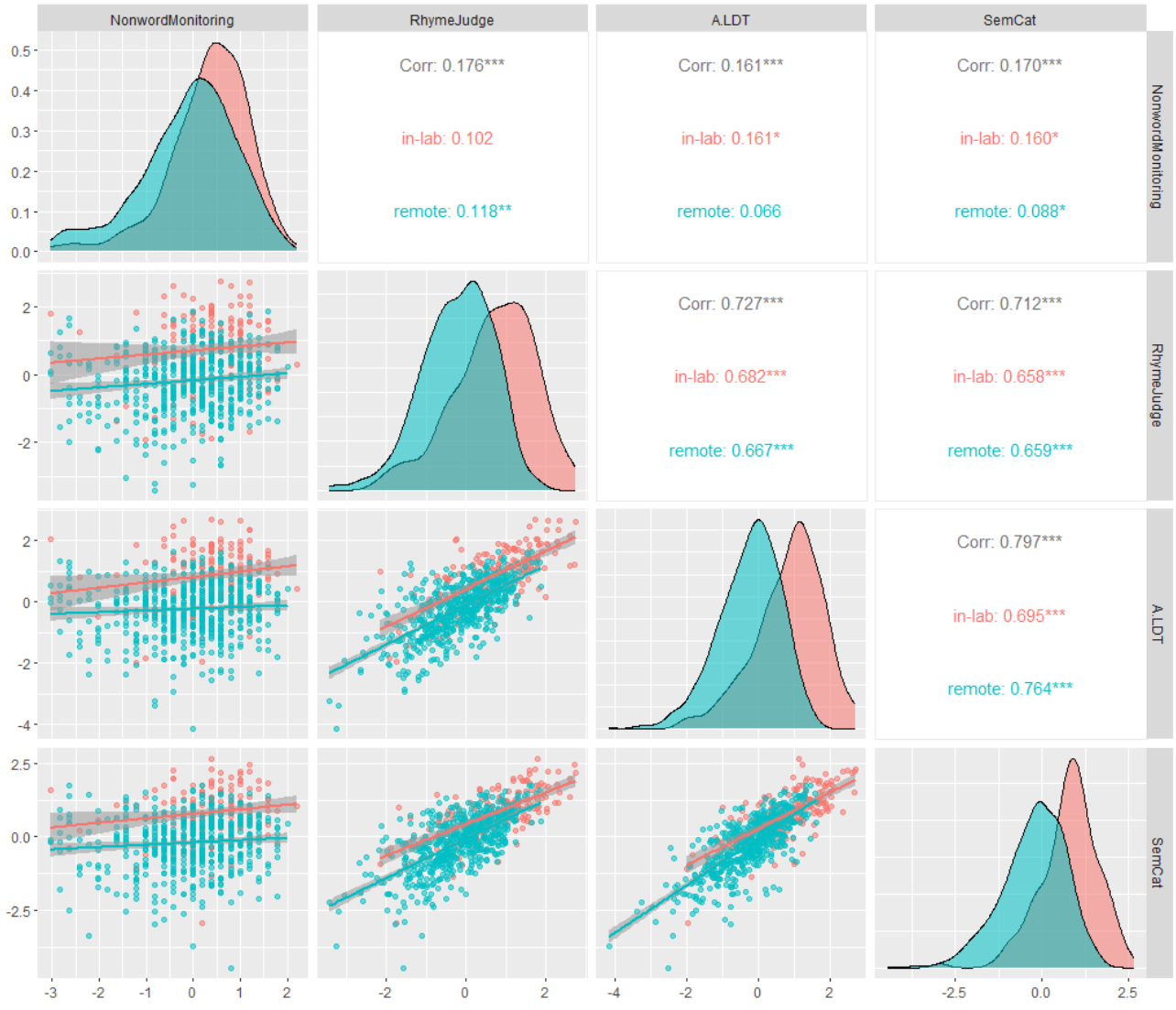

Figure 11

Score distributions for and correlations between *Sentence comprehension* (*Linguistic processing skills*) performance indicators for in-lab (red) and remote (green) experimental settings and for the combined data set.

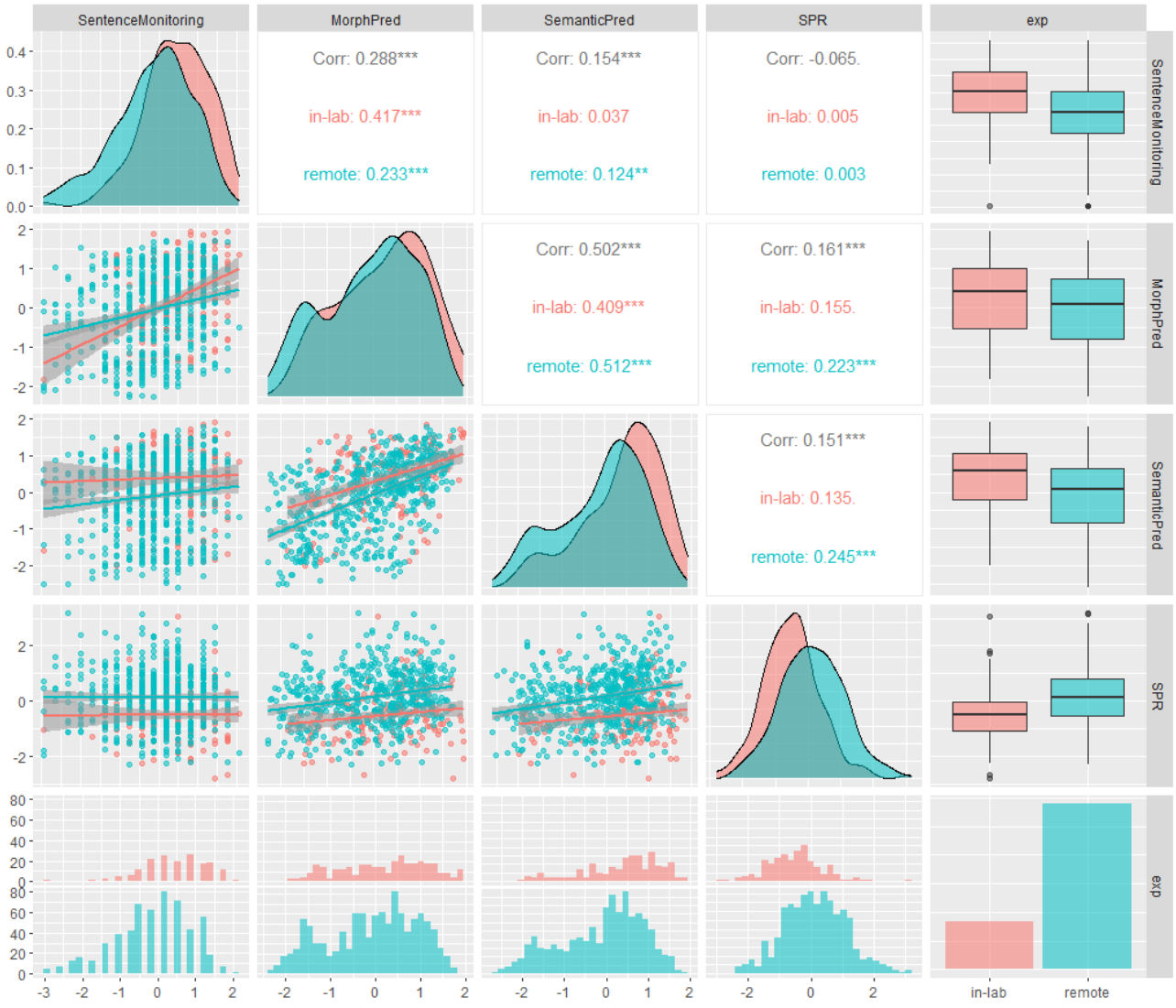

Table 12

Distribution and frequency plots for predicted *Linguistic processing skills* factor scores across in-lab (red) and remote (green) experimental settings.