

1 Comparison of ICCs using IRT and CTT parameters

2 Diego Figueiras¹ & John T. Kulas¹

3 ¹ Montclair State University

4 Author Note

5 Add complete departmental affiliations for each author here. Each new line herein
6 must be indented, like this line.

7 Enter author note here.

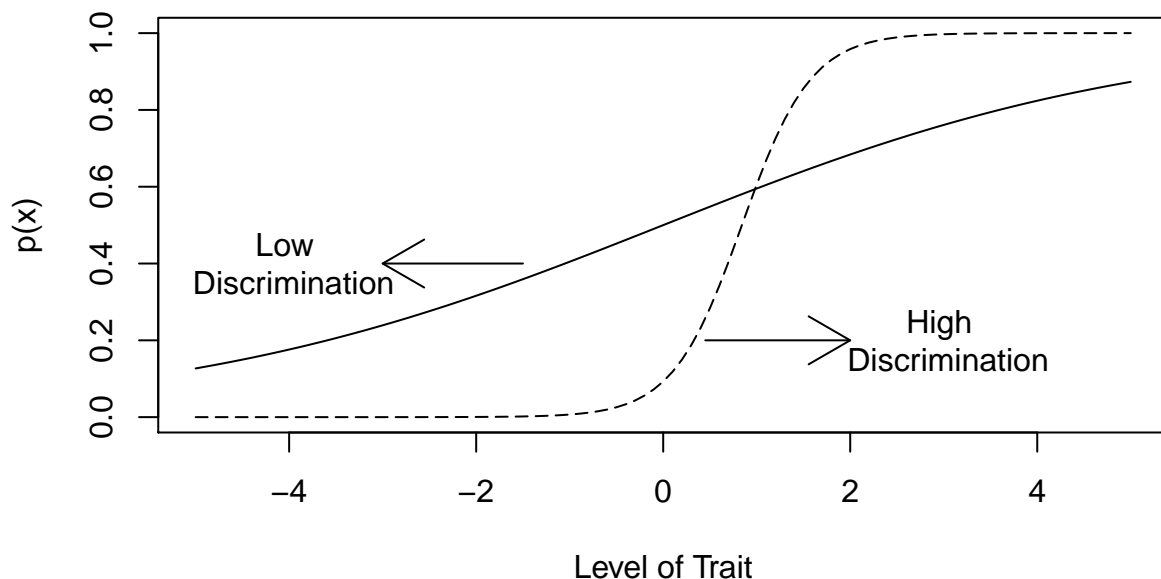
8 The authors made the following contributions. Diego Figueiras: Conceptualization,
9 Writing - Original Draft Preparation, Writing - Review & Editing; John T. Kulas: Writing
10 - Review & Editing.

11 Correspondence concerning this article should be addressed to Diego Figueiras,
12 Postal address. E-mail: figueirasd1@montclair.edu

Comparison of ICCs using IRT and CTT parameters

Introduction

Item characteristic curves are very often used by psychometricians to showcase and analyze the attributes of the item on a test or assessment. The x-axis shows a wide range of trait levels (ranging from high to low on the trait), while the y-axis displays probabilities of getting the item correct that range from 0 to 1. Each item has a curve. By looking at it, we can know the likelihood with which respondents of any trait level would answer any item correctly. If the curve is leaning towards the lower end of the trait level, this indicates that it is easy to answer the item correctly. On the contrary, if the curve is leaning towards the higher end of the trait level, this indicates that the item is difficult. If the curve is steep, this indicates high discrimination among respondents; if it is flat, it indicates no discrimination.

Item Characteristic Curves

Psychometricians who examine ICCs usually do it using Item Response Theory and Rasch models to get the parameters necessary to plot the curves. In a 2PL model, these would be item difficulty and item discrimination. Item difficulty is the necessary trait level for a respondent to have a 50/50 chance to answer the item correctly. Item discrimination is the degree to which an item can differentiate among individuals with low and high levels of the trait. From a Classical Test Theory (CTT) frame of thinking, the difficulty of an item is determined by looking at the p-values of the items, while discrimination is determined by checking the Cronbach alpha and the corrected item total correlations. Psychometricians who look at these CTT parameters don't typically use them to plot ICCs. There is no reason for them not to, since ICCs based on CTT parameters could provide information as valuable as those based on IRT or Rasch without the need of being familiar with these models and with how to compute the necessary estimates. Fan states in summary that IRT and CTT "... framework produce very similar item and person statistics" (p.379).

There is research that shows that there is little difference between the parameters of both frameworks. Lord (2012) described a function that approximates the relationship between IRT and CTT discrimination parameters. Although this wasn't intended for practical purposes but rather to assist in the conceptual comprehension of the discrimination parameter in IRT for people who were more familiar with CTT procedures, the formula was later modified by Kulas, Smith, and Xu (2017), with the purpose of minimizing the average residual. The formula is the following: [INSERT R EXPONENTIAL FORMULA]

Where r is the biserial corrected item total correlation of the item. Simulations identified systematic slope and inflection differences across item with differing b values, so the formula was further changed to include the following modifiers:

[INSERT FINAL FORMULA] Where g is the absolute deviation from 50% responding an item correctly and 50% responding incorrectly, and it's computed like this:

52 $g=|p-0.5|$. Z_g is the standard normal deviation associated with g . If we visualize the results
 53 of these re-specifications of Lord's formula using p-values (difficulty) of .5, .3 (or .7), and .1
 54 (or .9), and corrected item total correlations (discrimination) of .3, .7 and .1, respectively,
 55 we get the following:

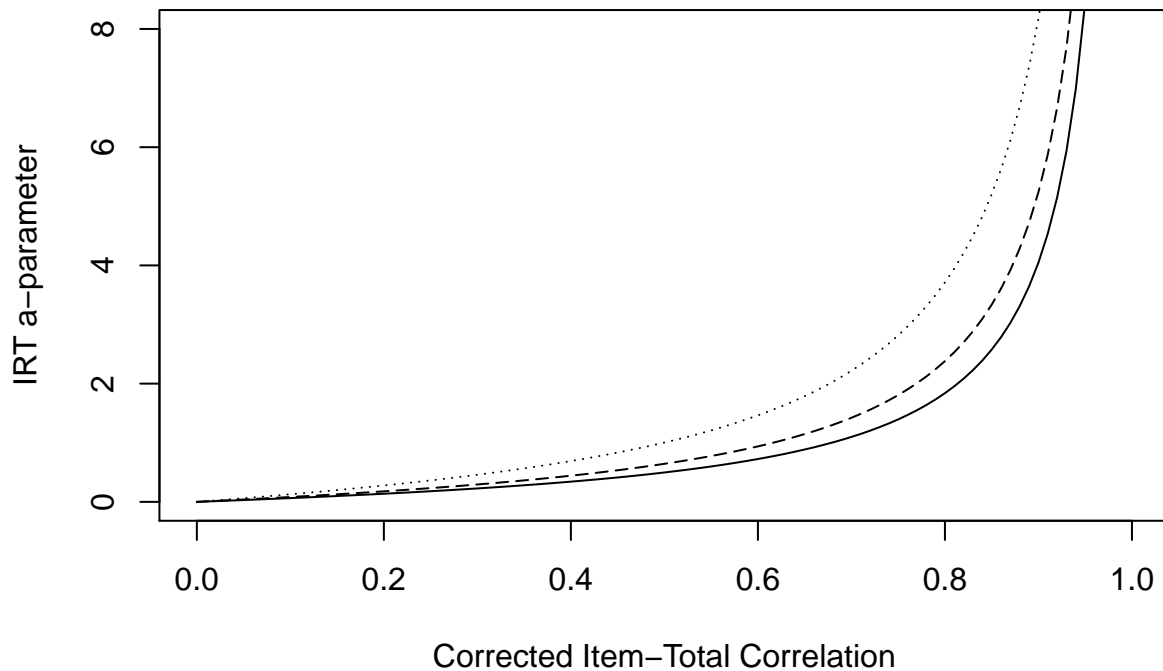


Figure 1. Functional relationship between the IRT a parameter and the CTT corrected-item total correlation as a function of item difficulty (p-value; solid = .5, dashed = .3/.7, dotted = .1/.9).

56 As we can see, the higher the corrected item-total correlations, the higher the
 57 estimated IRT a -parameter (discrimination). Also, as the p-values (difficulty) deviates
 58 from 0, the relationship between the estimated IRT a -parameter and the corrected
 59 item-total correlations becomes stronger.

60 Practitioners and researchers that don't use IRT or Rasch models and instead opt to

61 follow a CTT philosophy would benefit from having ICCs that use CTT statistics. This
62 study intends to show evidence of the overlapping nature of CTT and IRT parameters
63 when it comes to plotting ICCs.

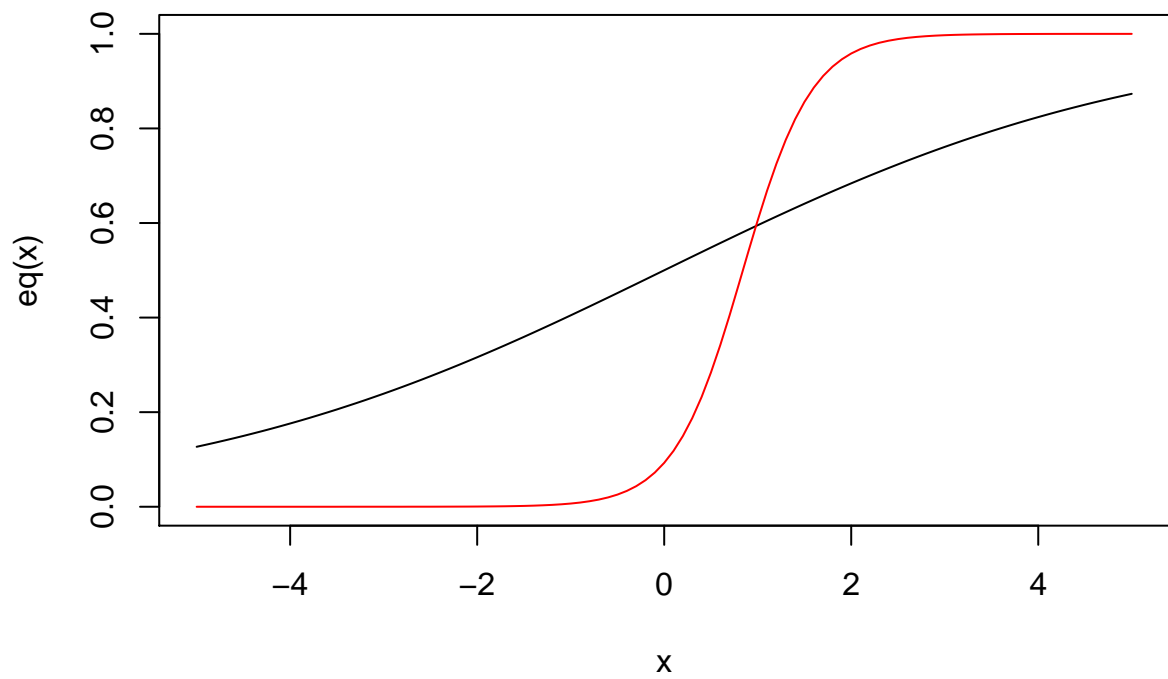
64 **Study 1 - Visual of discrimination relationship**

65 The purpose of study 1 is to look at the visualizations resulting from Kulas et al.
66 (2017) formula on simulated data. We hypothesize that the relationship between the
67 estimated IRT a-parameter and the corrected item-total correlations will be stronger as the
68 later deviates from 0, which would mean that the item has more discrimination.

69 **Procedure and methods**

70 We simulated data using Han (2007) software. Our sample was 10,000 observations,
71 with a mean of 0 and a standard deviation of 1. The number of itemm were 50, with
72 response categories of either correct or incorrect (1 and 0).

73 Study 2 simulates a bunch of test data and then we generate ICCs based on the IRT
74 model and then we compare that to our CTT estimates.

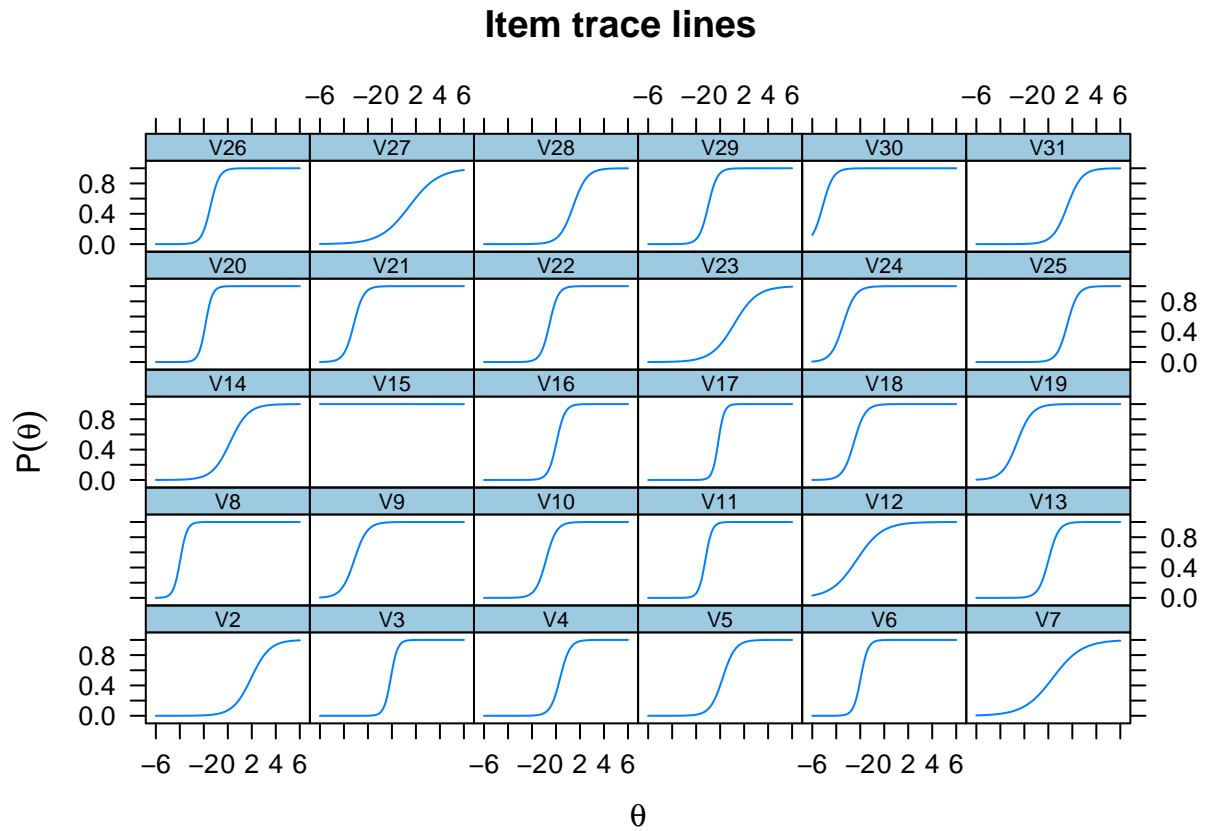


75

76 ## Iteration: 1, Log-Lik: -98116.710, Max-Change: 4.09843Iteration: 2, Log-Lik: -93555.3

77 ##

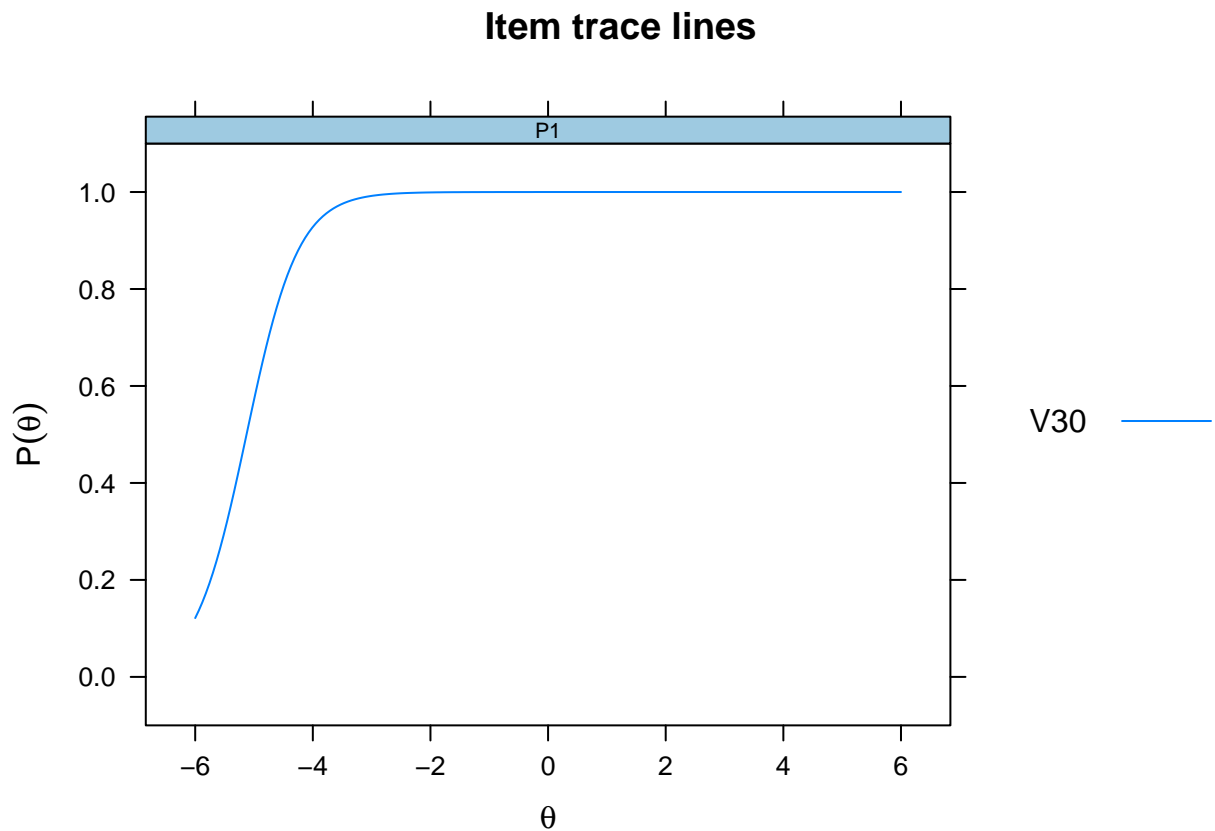
78 ## Calculating information matrix...



79

80

Iteration: 1, Log-Lik: -98116.710, Max-Change: 4.09843 Iteration: 2, Log-Lik: -93555.3



81

82 **##** Some items (V15) were negatively correlated with the total scale and

83 **##** probably should be reversed.

84 **##** To do this, run the function again with the 'check.keys=TRUE' option

85 **##**

86 **##** Reliability analysis

87 **##** Call: `alpha(x = data)`

88 **##**

89 **##** raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r

90 **##** 0.88 0.86 0.87 0.17 6.2 0.0015 0.65 0.17 0.15

91 **##**

92 **##** lower alpha upper 95% confidence boundaries

93 **##** 0.88 0.88 0.89

94 ##

95 ## Reliability if an item is dropped:

96 ##	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
97 ## V2	0.88	0.86	0.87	0.17	6.1	0.0015	0.020	0.16	
98 ## V3	0.87	0.85	0.86	0.16	5.7	0.0016	0.017	0.15	
99 ## V4	0.87	0.85	0.86	0.17	5.8	0.0016	0.018	0.15	
100 ## V5	0.88	0.85	0.87	0.17	5.8	0.0015	0.018	0.15	
101 ## V6	0.88	0.86	0.87	0.17	6.0	0.0015	0.019	0.15	
102 ## V7	0.88	0.86	0.87	0.17	6.1	0.0014	0.020	0.16	
103 ## V8	0.88	0.87	0.88	0.18	6.4	0.0015	0.018	0.17	
104 ## V9	0.88	0.86	0.87	0.18	6.2	0.0015	0.020	0.16	
105 ## V10	0.88	0.85	0.86	0.17	5.8	0.0015	0.018	0.15	
106 ## V11	0.88	0.85	0.86	0.17	5.8	0.0015	0.018	0.15	
107 ## V12	0.88	0.86	0.87	0.17	6.1	0.0015	0.020	0.16	
108 ## V13	0.87	0.85	0.86	0.17	5.8	0.0016	0.018	0.15	
109 ## V14	0.88	0.86	0.87	0.17	5.9	0.0015	0.019	0.15	
110 ## V15	0.88	0.87	0.88	0.18	6.5	0.0015	0.018	0.17	
111 ## V16	0.87	0.85	0.86	0.16	5.7	0.0016	0.017	0.15	
112 ## V17	0.87	0.85	0.86	0.16	5.6	0.0016	0.017	0.15	
113 ## V18	0.88	0.86	0.87	0.17	6.1	0.0015	0.020	0.16	
114 ## V19	0.88	0.86	0.87	0.17	6.1	0.0015	0.020	0.16	
115 ## V20	0.88	0.86	0.87	0.17	5.9	0.0015	0.019	0.15	
116 ## V21	0.88	0.86	0.87	0.18	6.3	0.0015	0.019	0.16	
117 ## V22	0.87	0.85	0.86	0.16	5.7	0.0016	0.018	0.15	
118 ## V23	0.88	0.86	0.87	0.17	6.1	0.0015	0.020	0.15	
119 ## V24	0.88	0.86	0.87	0.18	6.3	0.0015	0.019	0.16	
120 ## V25	0.88	0.86	0.87	0.17	6.1	0.0015	0.019	0.15	

```

121 ## V26      0.88      0.85      0.87      0.17 5.8      0.0015 0.019  0.15
122 ## V27      0.88      0.86      0.87      0.17 6.1      0.0014 0.020  0.16
123 ## V28      0.88      0.86      0.87      0.17 6.0      0.0015 0.019  0.15
124 ## V29      0.88      0.85      0.86      0.17 5.8      0.0015 0.018  0.15
125 ## V30      0.88      0.87      0.88      0.18 6.5      0.0015 0.018  0.17
126 ## V31      0.88      0.86      0.87      0.17 6.1      0.0015 0.019  0.15
127 ##
128 ## Item statistics
129 ##          n raw.r std.r  r.cor r.drop mean    sd
130 ## V2  10000  0.366  0.36  0.3121  0.3092  0.12  0.325
131 ## V3  10000  0.748  0.70  0.7108  0.6999  0.53  0.499
132 ## V4  10000  0.665  0.62  0.6141  0.6074  0.39  0.488
133 ## V5  10000  0.629  0.58  0.5730  0.5656  0.44  0.496
134 ## V6  10000  0.395  0.47  0.4512  0.3613  0.96  0.203
135 ## V7  10000  0.422  0.39  0.3453  0.3386  0.43  0.495
136 ## V8  10000  0.041  0.13  0.0656  0.0383  1.00  0.014
137 ## V9  10000  0.198  0.28  0.2306  0.1762  0.99  0.115
138 ## V10 10000  0.627  0.61  0.6077  0.5717  0.74  0.436
139 ## V11 10000  0.590  0.62  0.6151  0.5437  0.86  0.343
140 ## V12 10000  0.341  0.35  0.3060  0.2802  0.86  0.343
141 ## V13 10000  0.681  0.63  0.6307  0.6235  0.49  0.500
142 ## V14 10000  0.545  0.51  0.4800  0.4718  0.45  0.498
143 ## V15 10000 -0.003  0.07 -0.0044 -0.0049  1.00  0.010
144 ## V16 10000  0.716  0.67  0.6708  0.6633  0.48  0.500
145 ## V17 10000  0.763  0.71  0.7286  0.7179  0.56  0.497
146 ## V18 10000  0.291  0.36  0.3174  0.2612  0.97  0.164
147 ## V19 10000  0.303  0.36  0.3100  0.2668  0.96  0.201

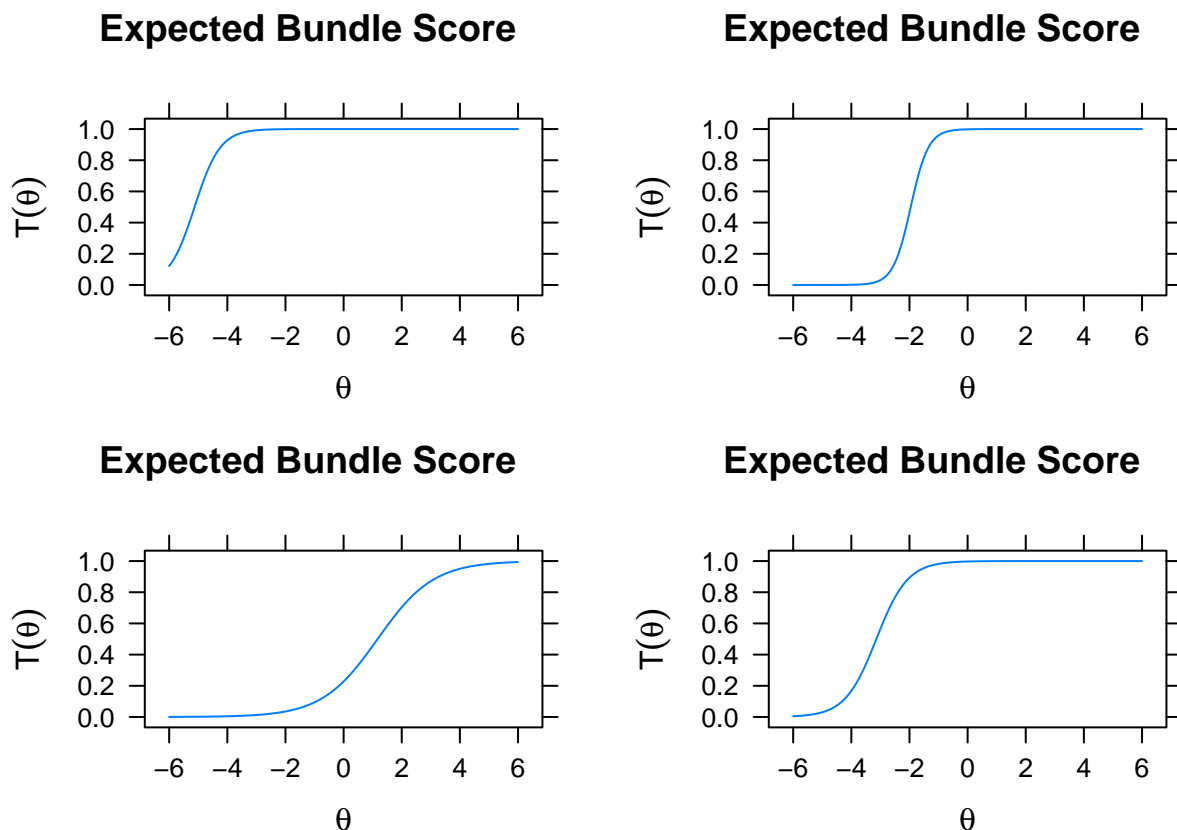
```

```

148 ## V20 10000 0.419 0.49 0.4751 0.3838 0.95 0.216
149 ## V21 10000 0.160 0.25 0.1894 0.1450 0.99 0.077
150 ## V22 10000 0.691 0.67 0.6672 0.6386 0.68 0.468
151 ## V23 10000 0.434 0.41 0.3677 0.3608 0.27 0.443
152 ## V24 10000 0.154 0.24 0.1777 0.1384 0.99 0.083
153 ## V25 10000 0.431 0.42 0.3818 0.3810 0.10 0.306
154 ## V26 10000 0.529 0.56 0.5525 0.4825 0.89 0.318
155 ## V27 10000 0.375 0.35 0.3054 0.2985 0.26 0.438
156 ## V28 10000 0.457 0.44 0.4022 0.3987 0.16 0.362
157 ## V29 10000 0.622 0.62 0.6200 0.5709 0.80 0.403
158 ## V30 10000 0.024 0.10 0.0314 0.0223 1.00 0.010
159 ## V31 10000 0.430 0.41 0.3762 0.3733 0.14 0.345
160 ##
161 ## Non missing response frequency for each item
162 ##      0      1 miss
163 ## V2  0.88 0.12    0
164 ## V3  0.47 0.53    0
165 ## V4  0.61 0.39    0
166 ## V5  0.56 0.44    0
167 ## V6  0.04 0.96    0
168 ## V7  0.57 0.43    0
169 ## V8  0.00 1.00    0
170 ## V9  0.01 0.99    0
171 ## V10 0.26 0.74    0
172 ## V11 0.14 0.86    0
173 ## V12 0.14 0.86    0
174 ## V13 0.51 0.49    0

```

175	##	V14	0.55	0.45	0
176	##	V15	0.00	1.00	0
177	##	V16	0.52	0.48	0
178	##	V17	0.44	0.56	0
179	##	V18	0.03	0.97	0
180	##	V19	0.04	0.96	0
181	##	V20	0.05	0.95	0
182	##	V21	0.01	0.99	0
183	##	V22	0.32	0.68	0
184	##	V23	0.73	0.27	0
185	##	V24	0.01	0.99	0
186	##	V25	0.90	0.10	0
187	##	V26	0.11	0.89	0
188	##	V27	0.74	0.26	0
189	##	V28	0.84	0.16	0
190	##	V29	0.20	0.80	0
191	##	V30	0.00	1.00	0
192	##	V31	0.86	0.14	0



193

194 ## Some items (V15) were negatively correlated with the total scale and

195 ## probably should be reversed.

196 ## To do this, run the function again with the 'check.keys=TRUE' option

197 ##

198 ## Reliability analysis

199 ## Call: alpha(x = data)

200 ##

201 ## raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r

202 ## 0.88 0.86 0.87 0.17 6.2 0.0015 0.65 0.17 0.15

203 ##

204 ## lower alpha upper 95% confidence boundaries

205 ## 0.88 0.88 0.89

206 ##

207 ## Reliability if an item is dropped:

208 ##	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
209 ## V2	0.88	0.86	0.87	0.17	6.1	0.0015	0.020	0.16	
210 ## V3	0.87	0.85	0.86	0.16	5.7	0.0016	0.017	0.15	
211 ## V4	0.87	0.85	0.86	0.17	5.8	0.0016	0.018	0.15	
212 ## V5	0.88	0.85	0.87	0.17	5.8	0.0015	0.018	0.15	
213 ## V6	0.88	0.86	0.87	0.17	6.0	0.0015	0.019	0.15	
214 ## V7	0.88	0.86	0.87	0.17	6.1	0.0014	0.020	0.16	
215 ## V8	0.88	0.87	0.88	0.18	6.4	0.0015	0.018	0.17	
216 ## V9	0.88	0.86	0.87	0.18	6.2	0.0015	0.020	0.16	
217 ## V10	0.88	0.85	0.86	0.17	5.8	0.0015	0.018	0.15	
218 ## V11	0.88	0.85	0.86	0.17	5.8	0.0015	0.018	0.15	
219 ## V12	0.88	0.86	0.87	0.17	6.1	0.0015	0.020	0.16	
220 ## V13	0.87	0.85	0.86	0.17	5.8	0.0016	0.018	0.15	
221 ## V14	0.88	0.86	0.87	0.17	5.9	0.0015	0.019	0.15	
222 ## V15	0.88	0.87	0.88	0.18	6.5	0.0015	0.018	0.17	
223 ## V16	0.87	0.85	0.86	0.16	5.7	0.0016	0.017	0.15	
224 ## V17	0.87	0.85	0.86	0.16	5.6	0.0016	0.017	0.15	
225 ## V18	0.88	0.86	0.87	0.17	6.1	0.0015	0.020	0.16	
226 ## V19	0.88	0.86	0.87	0.17	6.1	0.0015	0.020	0.16	
227 ## V20	0.88	0.86	0.87	0.17	5.9	0.0015	0.019	0.15	
228 ## V21	0.88	0.86	0.87	0.18	6.3	0.0015	0.019	0.16	
229 ## V22	0.87	0.85	0.86	0.16	5.7	0.0016	0.018	0.15	
230 ## V23	0.88	0.86	0.87	0.17	6.1	0.0015	0.020	0.15	
231 ## V24	0.88	0.86	0.87	0.18	6.3	0.0015	0.019	0.16	
232 ## V25	0.88	0.86	0.87	0.17	6.1	0.0015	0.019	0.15	

```

233 ## V26      0.88      0.85      0.87      0.17 5.8      0.0015 0.019  0.15
234 ## V27      0.88      0.86      0.87      0.17 6.1      0.0014 0.020  0.16
235 ## V28      0.88      0.86      0.87      0.17 6.0      0.0015 0.019  0.15
236 ## V29      0.88      0.85      0.86      0.17 5.8      0.0015 0.018  0.15
237 ## V30      0.88      0.87      0.88      0.18 6.5      0.0015 0.018  0.17
238 ## V31      0.88      0.86      0.87      0.17 6.1      0.0015 0.019  0.15
239 ##
240 ## Item statistics
241 ##          n raw.r std.r   r.cor r.drop mean    sd
242 ## V2  10000  0.366  0.36  0.3121  0.3092 0.12 0.325
243 ## V3  10000  0.748  0.70  0.7108  0.6999 0.53 0.499
244 ## V4  10000  0.665  0.62  0.6141  0.6074 0.39 0.488
245 ## V5  10000  0.629  0.58  0.5730  0.5656 0.44 0.496
246 ## V6  10000  0.395  0.47  0.4512  0.3613 0.96 0.203
247 ## V7  10000  0.422  0.39  0.3453  0.3386 0.43 0.495
248 ## V8  10000  0.041  0.13  0.0656  0.0383 1.00 0.014
249 ## V9  10000  0.198  0.28  0.2306  0.1762 0.99 0.115
250 ## V10 10000  0.627  0.61  0.6077  0.5717 0.74 0.436
251 ## V11 10000  0.590  0.62  0.6151  0.5437 0.86 0.343
252 ## V12 10000  0.341  0.35  0.3060  0.2802 0.86 0.343
253 ## V13 10000  0.681  0.63  0.6307  0.6235 0.49 0.500
254 ## V14 10000  0.545  0.51  0.4800  0.4718 0.45 0.498
255 ## V15 10000 -0.003  0.07 -0.0044 -0.0049 1.00 0.010
256 ## V16 10000  0.716  0.67  0.6708  0.6633 0.48 0.500
257 ## V17 10000  0.763  0.71  0.7286  0.7179 0.56 0.497
258 ## V18 10000  0.291  0.36  0.3174  0.2612 0.97 0.164
259 ## V19 10000  0.303  0.36  0.3100  0.2668 0.96 0.201

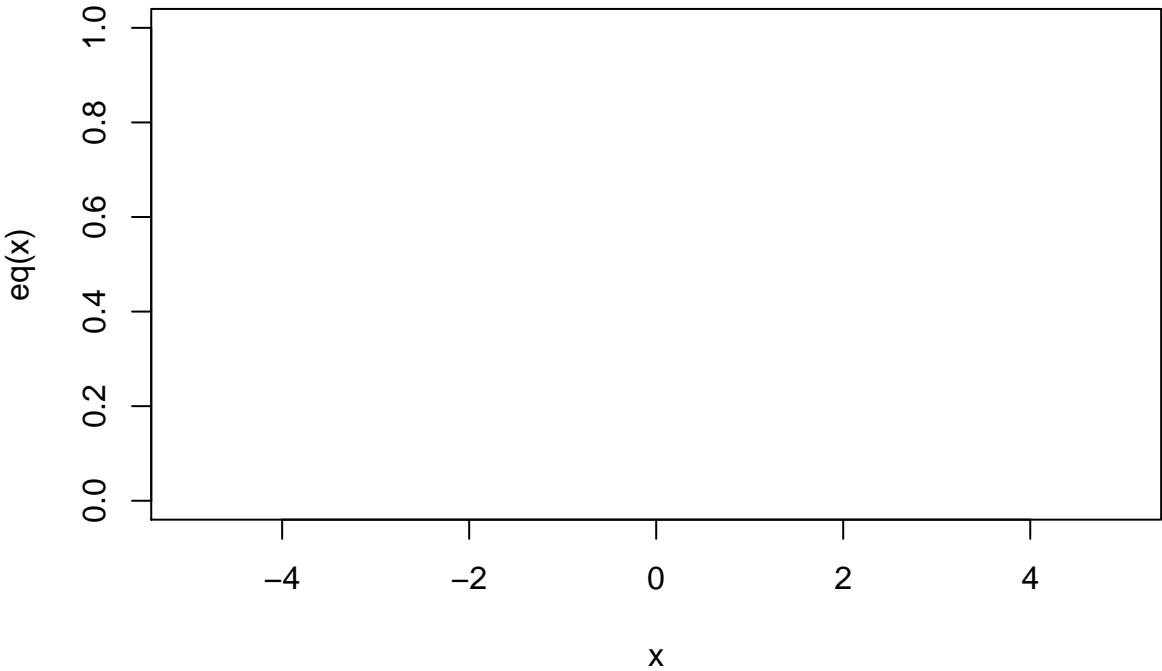
```

```

260 ## V20 10000 0.419 0.49 0.4751 0.3838 0.95 0.216
261 ## V21 10000 0.160 0.25 0.1894 0.1450 0.99 0.077
262 ## V22 10000 0.691 0.67 0.6672 0.6386 0.68 0.468
263 ## V23 10000 0.434 0.41 0.3677 0.3608 0.27 0.443
264 ## V24 10000 0.154 0.24 0.1777 0.1384 0.99 0.083
265 ## V25 10000 0.431 0.42 0.3818 0.3810 0.10 0.306
266 ## V26 10000 0.529 0.56 0.5525 0.4825 0.89 0.318
267 ## V27 10000 0.375 0.35 0.3054 0.2985 0.26 0.438
268 ## V28 10000 0.457 0.44 0.4022 0.3987 0.16 0.362
269 ## V29 10000 0.622 0.62 0.6200 0.5709 0.80 0.403
270 ## V30 10000 0.024 0.10 0.0314 0.0223 1.00 0.010
271 ## V31 10000 0.430 0.41 0.3762 0.3733 0.14 0.345
272 ##
273 ## Non missing response frequency for each item
274 ##      0      1 miss
275 ## V2  0.88 0.12    0
276 ## V3  0.47 0.53    0
277 ## V4  0.61 0.39    0
278 ## V5  0.56 0.44    0
279 ## V6  0.04 0.96    0
280 ## V7  0.57 0.43    0
281 ## V8  0.00 1.00    0
282 ## V9  0.01 0.99    0
283 ## V10 0.26 0.74    0
284 ## V11 0.14 0.86    0
285 ## V12 0.14 0.86    0
286 ## V13 0.51 0.49    0

```


287	##	V14	0.55	0.45	0
288	##	V15	0.00	1.00	0
289	##	V16	0.52	0.48	0
290	##	V17	0.44	0.56	0
291	##	V18	0.03	0.97	0
292	##	V19	0.04	0.96	0
293	##	V20	0.05	0.95	0
294	##	V21	0.01	0.99	0
295	##	V22	0.32	0.68	0
296	##	V23	0.73	0.27	0
297	##	V24	0.01	0.99	0
298	##	V25	0.90	0.10	0
299	##	V26	0.11	0.89	0
300	##	V27	0.74	0.26	0
301	##	V28	0.84	0.16	0
302	##	V29	0.20	0.80	0
303	##	V30	0.00	1.00	0
304	##	V31	0.86	0.14	0



305

306

307

Results

Discussion

References

- Han, K. (2007). WinGen3: Windows software that generates irt parameters and item responses [computer program]. *Amherst, MA: Center for Educational Assessment, University of Massachusetts Amherst.*
- Kulas, J. T., Smith, J. A., & Xu, H. (2017). Approximate functional relationship between irt and ctt item discrimination indices: A simulation, validation, and practical extension of lord's (1980) formula. *Journal of Applied Measurement, 18*(4), 393–407.
- Lord, F. M. (2012). *Applications of item response theory to practical testing problems.* Routledge.