Appendix A Analysis of ANES-EGSS4 Data

To show the generalizability of the proposed 2PL-G model, we also analyze the dataset from the American National Election Studies: Evaluations of Government and Society Study, Survey 4 (ANES-EGSS4) besides the TEDS data. There are 1,314 samples and four multiple-choice questions in the ANES-EGSS4 data: (1) Who is the Chief Justice of the U.S. Supreme Court (*Chief Justice*); (2) Who is the Prime Minister of the United Kingdom (*UK Premier*); (3) Who is the Speaker of the House of Representatives (*Speaker of HOR*); and (4) On which of the following does the U.S. federal government spend the least money (*Least Spending*).

Table A1: The Distributions of Choice Options against the Number of Item Success in ANES-EGSS4

Chief Justice No. of Success	John Roberts (Correct)	David Cole	Anthony Kennedy	Larry Thompson	No Answer
0	41.38 (72)	14.37 (25)	8.05 (14)	14.37 (25)	21.84 (38)
1	60.71 (221)	9.62 (35)	10.16 (37)	13.19 (48)	6.32 (23)
2	80.43 (333)	4.83 (20)	7.97 (33)	5.56 (23)	1.21 (5)
3	90.55 (278)	2.28 (7)	4.89 (15)	1.30 (4)	0.98 (3)
UK Premier No. of Success	David Cameron (Correct)	Nick Clegg	Tony Hayward	Richard Branson	No Answer
0	19.05 (24)	11.11 (14)	26.19 (33)	15.87 (20)	27.78 (35)
1	26.54 (69)	10.38 (27)	40.00 (104)	13.46 (35)	9.62 (25)
2	52.06 (253)	4.12 (20)	33.33 (162)	7.82 (38)	2.67 (13)
3	71.83 (278)	2.07 (8)	23.26 (90)	2.58 (10)	0.26 (1)
Speaker of HOR No. of Success	John Boehner (Correct)	Harry Reid	Eric Holder	Mitt Romney	No Answer
0	49.50 (100)	16.34 (33)	7.43 (15)	9.90 (20)	16.83 (34)
1	67.24 (236)	15.10 (53)	5.13 (18)	9.69 (34)	2.85 (10)
2	84.21 (352)	9.57 (40)	2.87 (12)	3.35 (14)	0.00 (0)
3	96.53 (278)	2.08 (6)	0.69 (2)	0.69 (2)	0.00 (0)
Least Spending No. of Success	Foreign Aid (Correct)	Medicare	National Defense	Social Security	No Answer
0	15.70 (19)	37.19 (45)	4.96 (6)	28.10 (34)	14.05 (17)
1	28.47 (78)	27.37 (75)	10.22 (28)	31.75 (87)	2.19 (6)
2	39.78 (148)	23.66 (88)	10.48 (39)	24.73 (92)	1.34 (5)
3	56.50 (278)	12.40 (61)	8.74 (43)	21.95 (108)	0.41 (2)

Note: Row percentages are presented and the corresponding counts are in parentheses; break-off and unit nonresponse are excluded.

Table A1 displays the distributions of choice options for the four items against the number of correct responses excluding the given item. We can see that Chief Justice and Speaker of HOR are

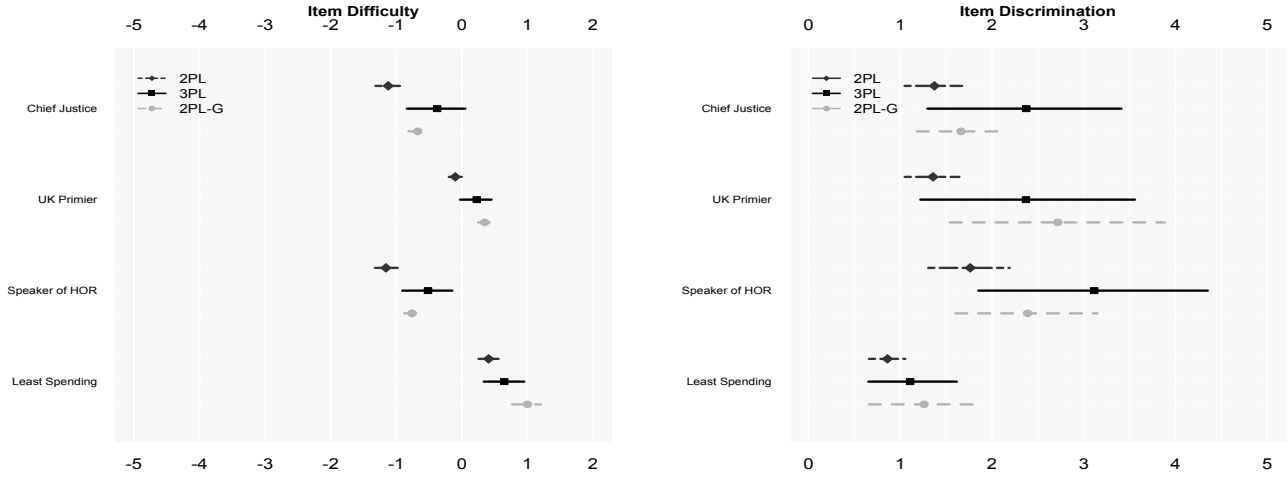

Figure A1: The 90% credible intervals of item parameters for the 2PL, 3PL, and 2PL-G models.

Figure A2: The 90% credible intervals of guessing effects for the 3PL and 2PL-G models.

relatively easy questions while UK Premier and Least Spending are relatively difficult questions. The distributions of choice options clearly demonstrate (mis)informed guessing rather than blind guessing in these four multiple-choice items, just like what we discuss in Section 4.1 for the Taiwan data.

We present the results of IRT analyses in Figures A1, A2, A3, and A4, which show the same patterns observed in the TEDS2012 data. In sum, first of all, the 2PL model has smaller values for the estimates of difficulty parameters, compared to the 3PL and 2PL-G models. Second, compared

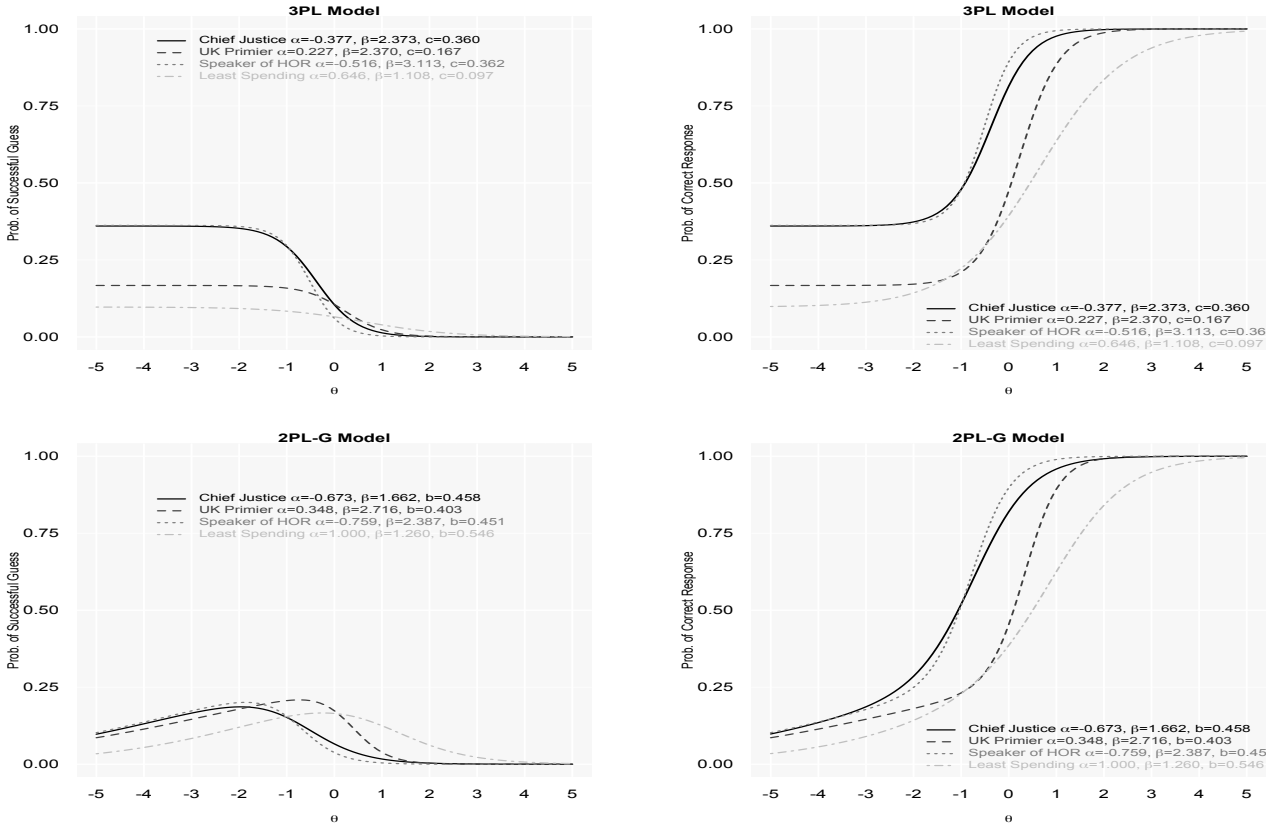

Figure A3: The probability of successful guesses and responses for the 3PL and 2PL-G models.

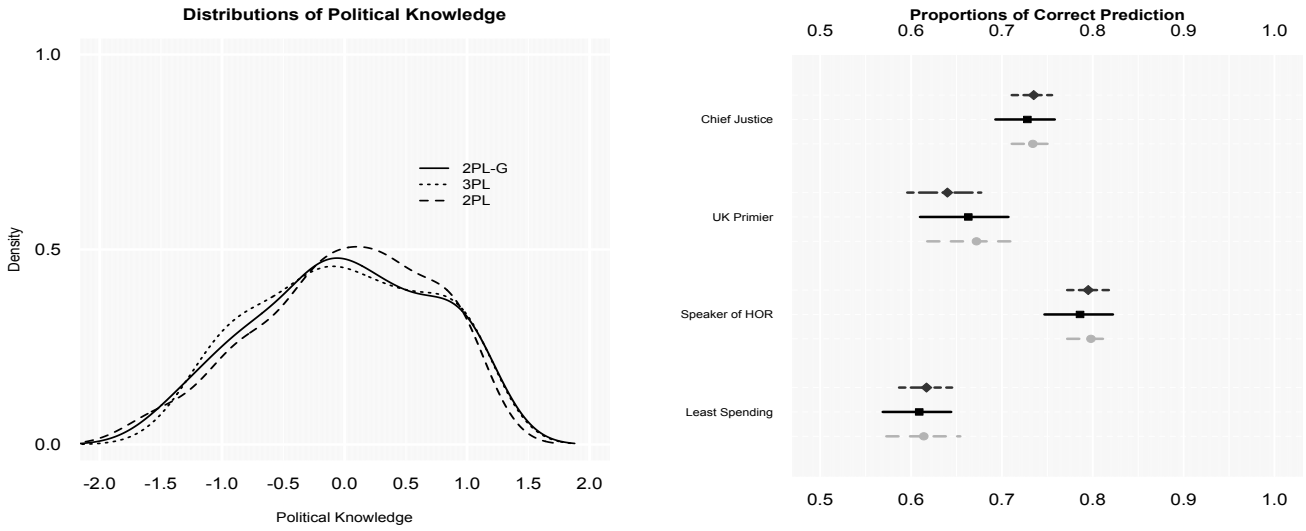

Figure A4: Comparisons between 2PL, 3PL, and 2PL-G. The left panel shows the distributions of political knowledge estimates. The right panel shows the 90% credible intervals of predictive accuracy.

to the 2PL-G model, the 3PL model has larger estimates of difficulty parameters for easy items while having smaller estimates of difficulty parameters for easy items due to the improperly high probability of successful guessing through parameter c. For easy items, the guessing component contributes too much to the probability of item success, which leads to a low probability of ability-based item success. Therefore, the difficulty parameters of easy items would be overestimated. By the same token, for hard questions, the guessing component does not contribute much to the probability of item success and, thus, a high probability of ability-based item success makes these items easier than they actually are.

Third, the two bottom panels of Figure A3 show that the 2PL-G model appropriately captures informed guessing for these four multiple-choice items. Regarding the 3PL model, in contrast, there is no reasonable explanation of the result that respondents with low knowledge levels have a higher probability of success than that of random guessing. Finally, in the right panel of Table A4 we present the predictive accuracy based on 879 respondents who have at least one correct response and at most three correct responses. There is no difference between the three models in predictive accuracy due to the overlapping of the 90% credible intervals. The reason why the three models do not differ in predictive accuracy might be that the range of estimated latent traits mostly lies between -1.5 and 1.5. In this range, the guessing components are not significantly different between the four items.

Appendix B Convergence Diagnostics for MCMC

As we mentioned in Footnote 15, we conducted diagnostic tests for the convergence of MCMC. To save space, we only present the results of item parameters for the 2PL-G model in this Appendix. The results of latent variables for the 2PL-G and the results of item parameters and latent variables for the 2PL and 3PL models can be obtained by running the replication files.

First, Table B1 shows the results of the Gelman-Rubin diagnostic. As can be seen, the potential scale reduction factors (PSRFs) of the 12 item parameters are close to 1 and the multivariate PSRF is close to 1 as well, which suggest the stability of Markov chains.

Table B1: Potential Scale Reduction Factors

Parameter	Point Est.	Upper C.I.
α_1	1.00	1.00
α_2	1.00	1.00
α_3	1.01	1.01
α_4	1.00	1.01
σ_α	1.00	1.00
β_1	1.00	1.00
β_2	1.00	1.00
β_3	1.01	1.02
β_4	1.00	1.01
b_1	1.00	1.00
b_2	1.00	1.00
b_3	1.00	1.00
b_4	1.00	1.00
Multivariate PSRF	1.01	

Second, Table B2 presents the results of the Geweke diagnostic. Based on an asymptotic standard normal statistic, the results suggest the stationarity because there are no appreciable differences in the means of the two periods.

Finally, Figure B1 displays the trace plots of samples from the three chains versus the iteration index for the item parameters b_k, which is assumed to follow a uniform distribution between 0 and 1. These trace plots show a mix of the three chains, which suggests the stationarity.

Table B2: Z-Scores from the Geweke Diagnostic

Parameter	$Chain1$	$Chain2$	$Chain3$
α_1	1.188	0.489	0.519
α_2	0.250	−0.390	−1.854
α_3	1.193	0.157	−1.735
α_4	0.206	−0.897	−0.748
σ_α	−0.705	−0.854	1.175
β_1	−1.401	−0.454	−0.362
β_2	0.670	0.278	1.380
β_3	0.858	−0.192	−1.850
β_4	−0.010	0.081	0.014
b_1	−0.577	0.079	−0.599
b_2	−1.766	1.066	0.107
b_3	−1.652	0.724	0.509
b_4	−0.021	1.345	0.524

Note: The first 10% and the last 50% of the values are compared.

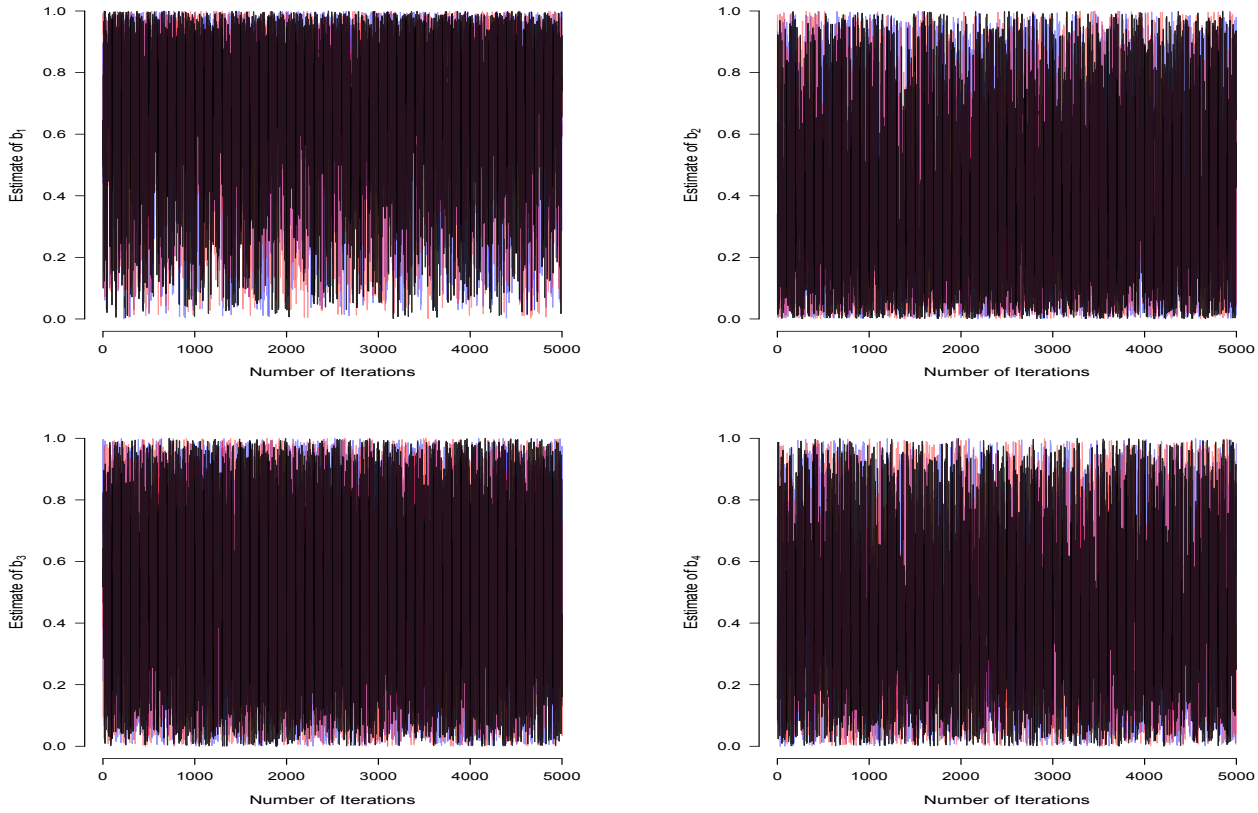

Figure B1: Trace Plots for Item Parameter b_k.

Appendix C Estimates of Parameter b_k of the 2PL-G Model

The means of parameter b_k for the four multiple-choice items are 0.669, 0.394, 0.499, and 0.436, respectively. The standard deviations are 0.248, 0.281, 0.289, and 0.262, respectively. Figure C1 displays the 90% credible intervals of these four parameters. Because the parameter b_k alone could not present the guessing property of multiple-choice items, we show the multiplication of b_k and α_k in Figure 4.

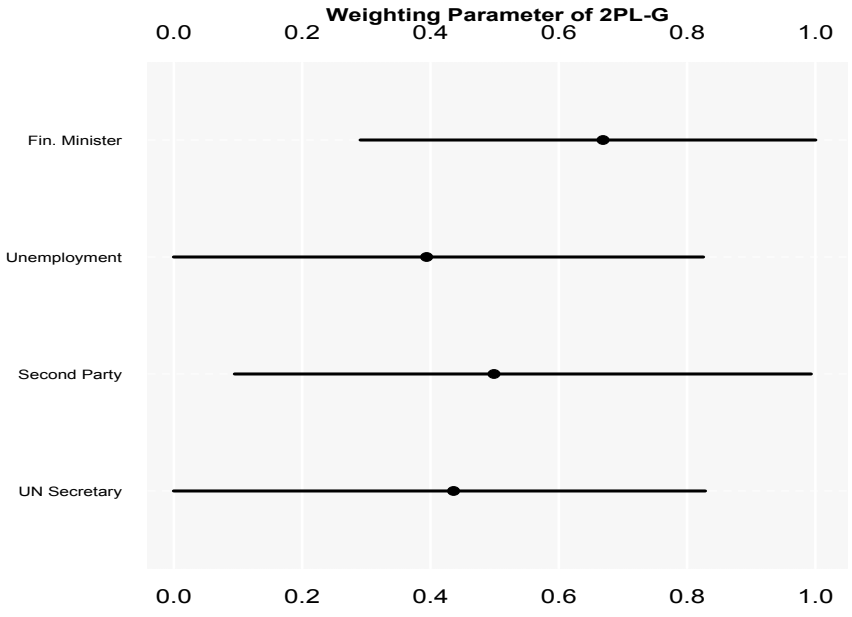

Figure C1: The 90% credible intervals of b_k for the four multiple-choice items.

Appendix D Comparisons between the Estimates of Political Knowledge from Different Models

Figure D1 shows the comparisons of political knowledge estimates from the 2PL, 3PL, and 2PL-G models. Because the 2PL model underestimates the difficulty parameters, it would overestimate political knowledge levels for the respondents with moderate and low levels of political knowledge, compared to the 2PL-G model. This phenomenon can be clearly seen in the left panel of Figure D1. Moreover, compared to the 2PL-G model, the 3PL model might overestimate difficulty parameters for easy items and the knowledge levels of barely informed respondents, which is displayed in the right panel of Figure D1.

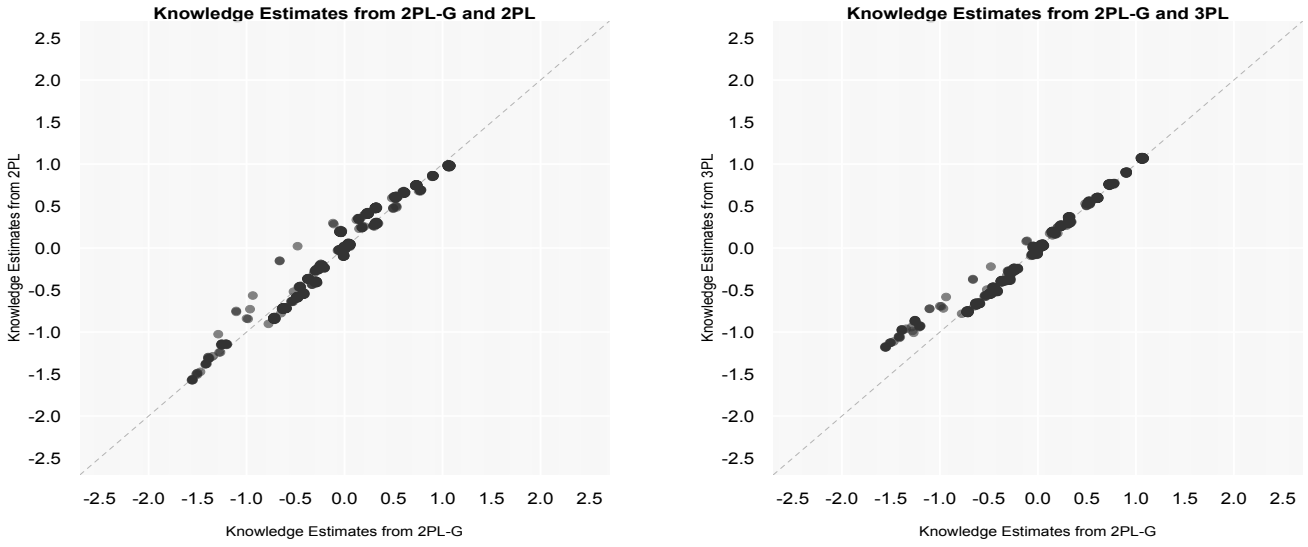

Figure D1: Comparisons between the Estimates of Political Knowledge from the 2PL, 3PL, and 2PL-G models.

Appendix E Constrained IRT Models

Although we observe some differences between the 2PL, 3PL, and 2PL-G models in Figure 6, they are not significantly different from one another. Here we add restrictions to item parameters and estimate the three models to see whether these differences are evident. We select the easiest item, Second Legislative Party, and fix its item-difficulty parameter to -1 and the item-discrimination parameter to 1.

The left panel of Figure E1 shows the distributions of estimated knowledge levels. We can see that the overestimation of latent traits by the 2PL model and the underestimation of latent traits by the 3PL mode are evident. Moreover, we observe the difference in the prediction accuracy for Second Legislative Party. The 2PL model is about 90% accurate while the 3PL and 2PL-G models are more than 95% accurate.

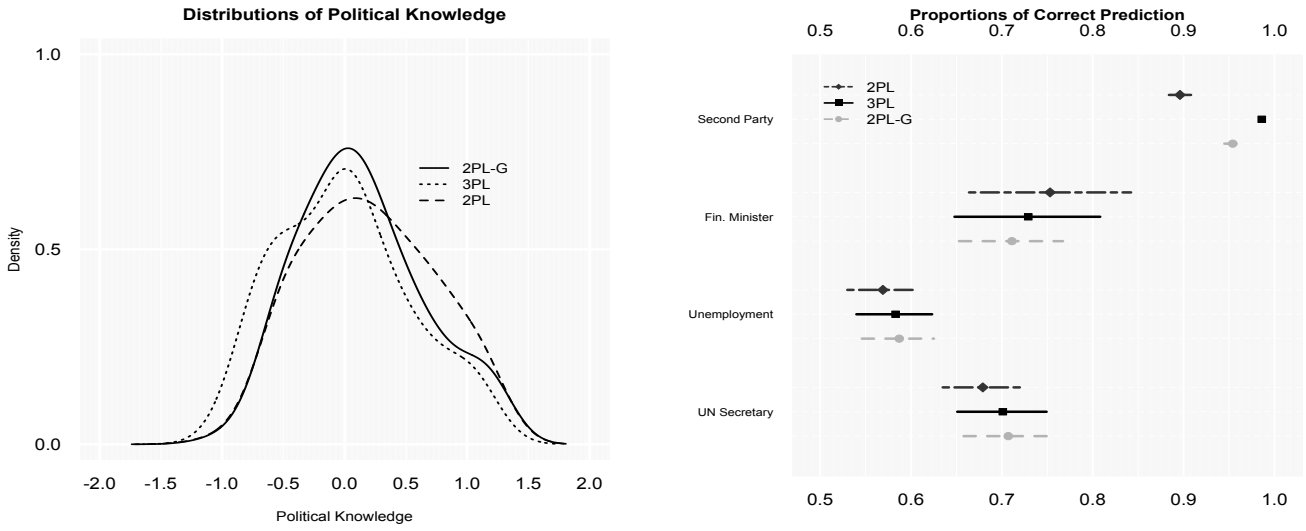

Figure E1: Comparisons between 2PL, 3PL, and 2PL-G. The left panel shows the distributions of political knowledge estimates. The right panel shows the 90% credible intervals of the proportions of correct predictions.

We also do the same restrictions to the other three items respectively. The results show that, when the item-difficulty and the item-discrimination parameters are fixed, the 2PL-G model performs better in predictive accuracy than the 2PL and 3PL models.

Appendix F Distributions of Choice Options across Gender in TEDS2012

Table F1: The Distributions of Choice Options against the Number of Item Success in TEDS 2012 for Female

Fin. Minister No. of Success	Jiang Yi-huah	Chen Chun	Mao Chi-kuo	Lee Sush-der (Correct)	"Don't Know"
0-1	6.37 (13)	5.88 (12)	2.45 (5)	2.45 (5)	82.84 (169)
2-4	11.58 (69)	21.14 (126)	9.23 (55)	26.68 (159)	31.38 (187)
5-6	11.43 (12)	16.19 (17)	3.81 (4)	60.95 (64)	7.62 (8)
Unemployment No. of Success	2.3%	4.3% (Correct)	6.3%	8.3%	"Don't Know"
0-1	2.39 (5)	5.26 (11)	9.57 (20)	10.05 (21)	72.73 (152)
2-4	2.21 (13)	30.90 (182)	30.73 (181)	7.64 (45)	28.52 (168)
5-6	0.93 (1)	50.47 (54)	30.84 (33)	8.41 (9)	9.35 (10)
Second Party No. of Success	KMT	DPP (Correct)	PFP	Non-Partisan Solidarity Union	"Don't Know"
0-1	3.85 (12)	61.86 (193)	0.00 (0)	0.32 (1)	33.97 (106)
2-4	1.72 (9)	94.85 (497)	0.38 (2)	0.19 (1)	2.86 (15)
5-6	1.45 (1)	98.55 (68)	0.00 (0)	0.00 (0)	0.00 (0)
UN Secretary No. of Success	Kofi Annan	Kurt Waldheim	Ban Ki-Moon (Correct)	Boutros Boutros-Ghali	"Don't Know"
0-1	1.44 (3)	4.33 (9)	4.81 (10)	1.44 (3)	87.98 (183)
2-4	19.89 (112)	5.68 (32)	12.08 (68)	2.84 (16)	59.50 (335)
5-6	34.33 (46)	1.49 (2)	39.55 (53)	0.75 (1)	23.88 (32)

Note: Row percentages are presented and the corresponding counts are in parentheses.

Table F2: The Distributions of Choice Options against the Number of Item Success in TEDS 2012 for Male

Fin. Minister No. of Success	Jiang Yi-huah	Chen Chun	Mao Chi-kuo	Lee Sush-der (Correct)	"Don't Know"
0-1	8.25 (8)	7.22 (7)	5.15 (5)	7.22 (7)	72.16 (70)
2-4	10.56 (62)	13.97 (82)	9.20 (54)	36.29 (213)	29.98 (176)
5-6	5.91 (14)	10.55 (25)	1.27 (3)	77.22 (183)	5.06 (12)
Unemployment No. of Success	2.3%	4.3% (Correct)	6.3%	8.3%	"Don't Know"
0-1	0.90 (1)	19.82 (22)	3.60 (4)	10.81 (12)	64.86 (72)
2-4	1.31 (7)	33.02 (177)	25.00 (134)	6.53 (35)	34.14 (183)
5-6	1.46 (4)	61.68 (169)	22.63 (62)	5.47 (15)	8.76 (24)
Second Party No. of Success	KMT	DPP (Correct)	PFP	Non-Partisan Solidarity Union	"Don't Know"
0-1	5.24 (11)	69.05 (145)	1.90 (4)	0.48 (1)	23.33 (49)
2-4	1.55 (8)	95.35 (492)	0.58 (3)	0.00 (0)	2.52 (13)
5-6	0.51 (1)	99.49 (194)	0.00 (0)	0.00 (0)	0.00 (0)
UN Secretary No. of Success	Kofi Annan	Kurt Waldheim	Ban Ki-Moon (Correct)	Boutros Boutros-Ghali	"Don't Know"
0-1	4.21 (4)	4.21 (4)	3.16 (3)	0.00 (0)	88.42 (84)
2-4	22.12 (115)	6.35 (33)	13.46 (70)	0.77 (4)	57.31 (298)
5-6	34.31 (105)	1.63 (5)	44.77 (137)	0.65 (2)	18.63 (57)

Note: Row percentages are presented and the corresponding counts are in parentheses.

Appendix G The Application of San Martín et al.'s Model

We apply the one-parameter logistic with ability-based guessing (1PL-AG) model developed by **?** to both the TEDS2012 data and the ANES-EGSS4 data. For the TEDS2012 data, the estimate of b is 0, which suggests that guessing correctly is not ability related. However, the positive value of guessing parameter γ for Second Party indicates that the effect of guessing is higher than average ability, which is not consistent to the estimate of $b = 0$. For the ANES-EGSS4 data, the estimate of b has mean 3.746 with the 90% credible interval $[0.927, 8.334]$.

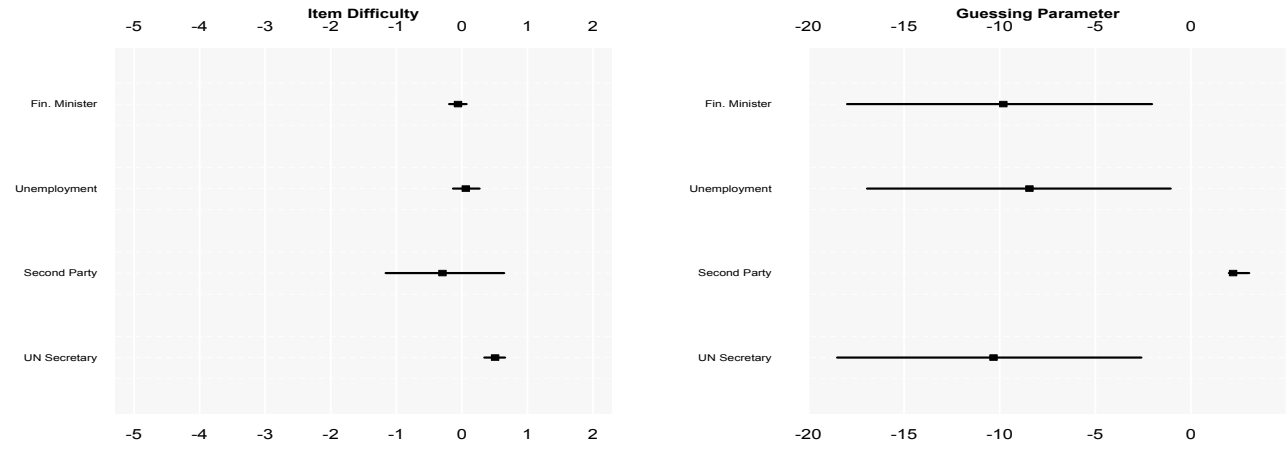

Figure G1: The 90% credible intervals of item parameters from TEDS2012 data.

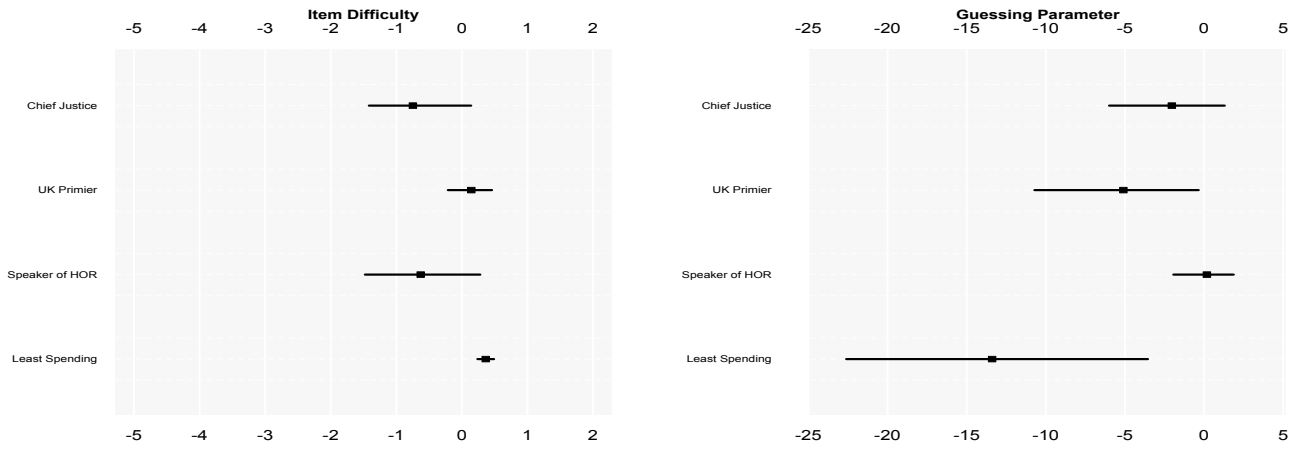

Figure G2: The 90% credible intervals of item parameters from ANES-EGSS4 data.