Supplementary Material

Items from a True-False Algebra Test and Answer Key

1)	$(a+b)\times c=(a\times c)+(b\times c)$	True	False
2)	$\frac{a}{c}\times b=\frac{b}{c}\times a$	True	False
3)	$a+(b\times c)=(a+b)\times(a+c)$	True	False
4)	$\frac{a\times(b+c)}{b}=a+\frac{a\times c}{b}$	True	False
5)	$\frac{a\times b}{c}=\frac{a}{c}\times\frac{b}{c}$	True	False
6)	$\frac{a/b}{c/d}=\frac{a\times c}{b\times d}$	True	False
7)	$\frac{1}{1/a}=a$	True	False
8)	$\sqrt{a+b}=\sqrt{a}+\sqrt{b}$	True	False
9)	$\frac{c}{a+b}=\frac{c}{a}+\frac{c}{b}$	True	False
10)	$\frac{1/a}{(1/b)^2}=\frac{b^2}{a}$	True	False
11)	$(a+b)^2=a^2+b^2$	True	False
12)	$\sqrt{a\times b}=\sqrt{a}\times\sqrt{b}$	True	False
13)	$(a+c)\times(b+d)=a\times(b+d)+c\times(b+d)$	True	False
14)	$\frac{a\times(b+c)^2}{1/a}=b^2+c^2+2(b\times c)$	True	False
15)	$\frac{1/a}{b}=\frac{b}{a}$	True	False
16)	$\frac{(a+b)^2}{b}=\frac{a^2}{b}+2a+b$	True	False

Note: the items were developed by the author.

Figure A1. Posterior predictive checks for the total test score, IRSDT and IRT models.

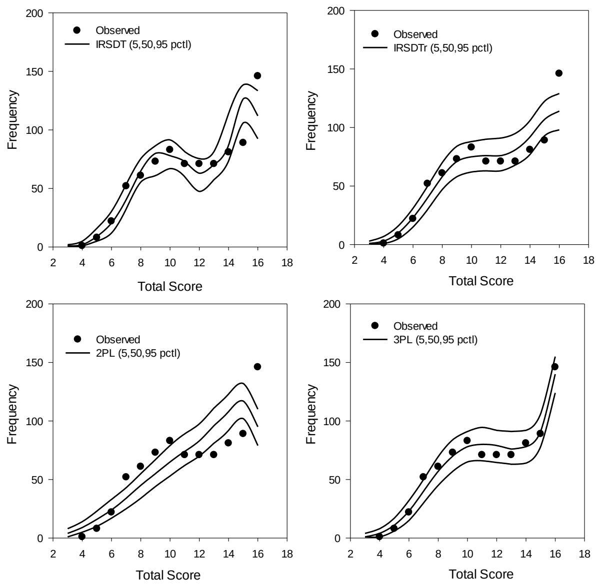
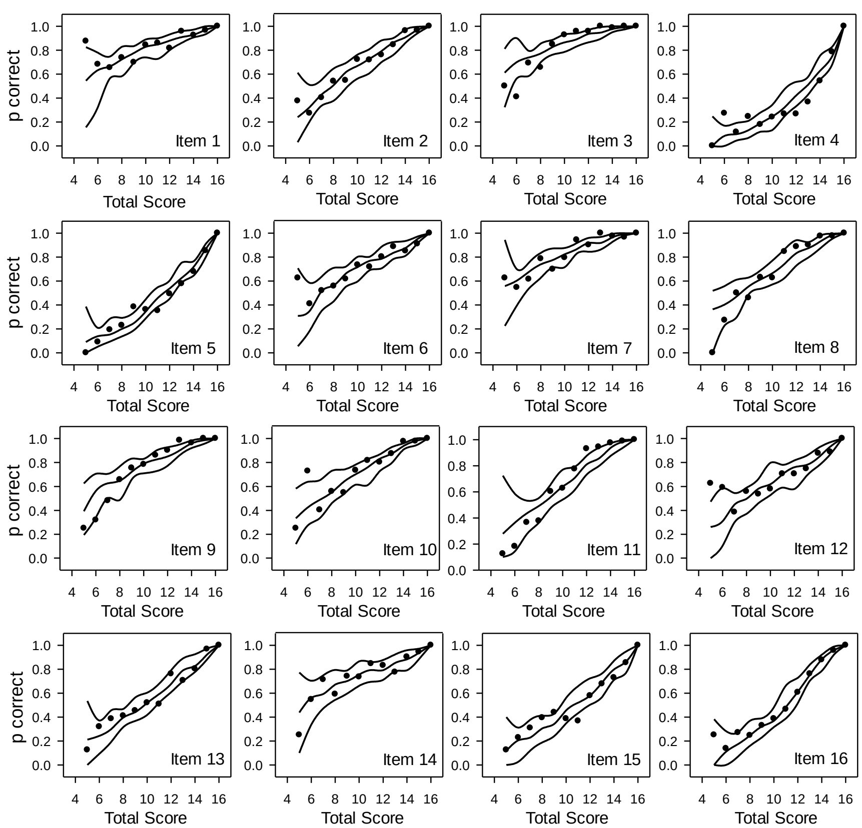


Figure A2. Posterior predictive checks for item-fit plots, IRSDT model.



A model with Varying Bias and Detection across Examinees

Several simplifying assumptions are made for the basic IRSDT model. One restrictive assumption, from an SDT point of view, is that the bias and detection parameters, b_j and d_j, are treated solely as *item characteristics* and so are assumed to be constant across the examinees. However, from the viewpoint of traditional SDT, these parameters might vary over examinees, reflecting individual differences in bias and detection. This suggests treating the item parameters as being *random* over examinees, rather than fixed, referred to here as the IRSDTr model. This can be accomplished by introducing additional examinee random variables, one for bias and one for detection, to reflect individual differences. Here it is assumed that, for a given examinee, the effects are the same across the items, which leads to a simple generalization that only adds two parameters. Thus, the fixed parameters are replaced by $b_j' = b_j + e_i$ with $e_i \sim N(0, \sigma_b^2)$ and $d_j' = d_j + f_i$ with $f_i \sim N(0, \sigma_d^2)$, and so b_j' and d_j' vary over examinees with variance σ_b^2 and σ_d^2, respectively, thus adding two parameters to the model. This means, for example, that if an examinee has higher detection compared to other examinees, detection is higher by the same amount across the j items for that examinee.

Although the basic and extended models can easily be implemented in current software, it should be noted that further work on identification and estimation issues needs to be done. The relation between the basic IRSDT model and the GoM model is important in this regards because it shows that work on these issues in the GoM literature is directly applicable to the IRSDT model, and so the model fits into an established framework. In addition, simulations presented next indicate acceptable estimation at least for the basic IRSDT model with a specified distribution for λ_i and (possibly) for the model with a beta distribution for λ_i with unknown shape parameters. Some small simulations with bimodal, Beta(.5, .5), uniform, Beta(1,1), and approaching normal, Beta(3,3), distributions suggested that one could detect the shape to some degree from the posteriors for ν and ω, however the posterior standard deviations were large and coverage was lower, and more work clearly needs to be done.

Simulations

Simulations were conducted to obtain some information about parameter recovery for the IRSDT and IRSDTr models. The simulated data were generated to be similar to the that found for the Algebra data and consisted of 16 items with a sample size of 500. Each condition consisted of 100 replications (i.e., 100 datasets for each condition). Population values of the parameters were chosen in line with those found for the Algebra data and are shown in Table A1; the bias b_j ranged from −2 to 2 and discrimination d_j ranged from 1 to 4. For each replication, Bayesian estimation of Equation 4 was done using PROC MCMC of SAS with 20,000 burn-ins followed by 100,000 iterations. The priors and hyperpriors used were: d ~ lognormal(0, 1) and b ~ N(0, 9), with λ_i ~ Beta(v,w) with v and w ~ lognormal(0,1), and for the IRSDTr model, $e_i \sim N(0, \sigma_b^2)$ and $f_i \sim N(0, \sigma_d^2)$ with σ_b^2 ~ lognormal(0,1) and σ_d^2 ~ lognormal(0,1). The models were examined first with a uniform distribution for λ_i, Beta(1,1), but other distributions – a bimodal distribution, Beta(.3,.3) and a more normal-like distribution, Beta(3,3) – were also examined.

Table A1 shows, for simulations for both the IRSDT and IRSDTr models with uniform lambda, the average posterior means, average posterior standard deviations, and bias (estimate minus population parameter) for each parameter over 100 replications, along with the obtained coverage rates for 90% confidence intervals. The first eight bias parameters all have population values of zero, and Table A1 shows that the estimates are all close to zero. The next eight bias parameters have values ranging from −2 to 2 and Table A1 shows that the estimates correctly indicate the size and direction of the population values.

With respect to the discrimination parameters, Table A1 shows that, for the IRSDT model, the posterior means show mostly negative positive bias and so discrimination tends to be underestimated, although the item discrimination magnitudes are still rank-ordered correctly. Table A1 shows that the obtained coverage rates for the IRSDT model parameters are generally close to the nominal 90% value, however the coverage rates are lower for the IRSDTr model. The bias for the shape parameters is small and negative and the coverage rates are over 80%. Recovery of the discrimination and bias variance parameters for the IRSDTr model are reasonable.

Table A2 shows the results for the conditions with a bimodal distribution. The IRSDT coverage rates are again generally close to the nominal 90% value, with the coverage again lower for the IRSDTr model. For the estimated shape parameters, the bias is large and positive and the coverage is poor, nevertheless the estimates indicate that the Beta is bimodal, in that they are both considerably less than one. Thus, the estimates provide some information about the general shape of the distribution, but not accurate estimates of the shape parameters; more research on estimation with known and unknown shape parameters and other options (nonparameteric) is needed.

Overall, the simulations suggest that the model parameter estimates provide useful information about the relative magnitudes of item bias and item discrimination. The shape parameters are acceptably recovered for uniform (and normal-like) situations, but less well recovered in the bimodal situation. However, for the uniform, normal, and bimodal situations, the estimates were around 1, clearly greater than 1, and less than 1, respectively, and so the estimates appear to provide some information about the shape of the λ_i distribution.

Table A1

Parameter Recovery for Simulated Data, Uniform λ_i, N=500

		IRSDT				IRSDTr			
Par.	Pop.	Av. PM	Av. PSD	Bias	% Cov.	Av. PM	Av. PSD	Bias	% Cov.
b_1	0	−0.12	0.22	−0.12	88	−0.25	0.25	−0.25	66
b_2	0	0.13	0.21	0.13	87	0.26	0.25	0.26	55
b_3	0	−0.10	0.22	−0.10	88	−0.21	0.25	−0.21	74
b_4	0	0.07	0.22	0.07	93	0.27	0.25	0.27	67
b_5	0	−0.09	0.22	−0.09	87	0.00	0.24	0.00	75
b_6	0	0.04	0.22	0.04	90	0.04	0.25	0.04	81
b_7	0	−0.03	0.20	−0.03	98	0.10	0.23	0.10	86
b_8	0	0.02	0.20	0.02	96	−0.15	0.25	−0.15	88
b_9	1	0.97	0.25	−0.03	91	1.06	0.24	0.06	84
b_{10}	−1	−0.97	0.27	0.02	93	−1.07	0.23	−0.07	89
b_{11}	1.5	1.38	0.37	−0.12	85	1.43	0.31	−0.07	84
b_{12}	−1.5	−1.38	0.36	0.12	89	−1.38	0.30	0.12	74
b_{13}	2	1.83	0.52	−0.17	82	1.64	0.35	−0.36	64
b_{14}	−2	−1.87	0.50	0.13	90	−1.63	0.34	0.37	51
b_{15}	0.5	0.41	0.27	−0.09	86	0.16	0.26	−0.34	57
b_{16}	−0.5	−0.39	0.26	0.11	89	−0.17	0.26	0.33	58
d_1	4	3.84	1.15	−0.16	85	2.19	0.57	−1.81	28
d_2	4	3.63	1.04	−0.37	77	2.11	0.60	−1.89	25
d_3	3	2.99	0.93	−0.01	82	1.82	0.55	−1.18	39
d_4	3	3.05	0.97	0.05	85	1.68	0.55	−1.32	37
d_5	2	1.94	0.68	−0.06	89	1.65	0.52	−0.35	59
d_6	2	2.02	0.68	0.02	92	1.60	0.55	−0.40	67
d_7	1	0.96	0.38	−0.04	97	1.04	0.46	0.04	83
d_8	1	0.98	0.40	−0.19	94	1.19	0.51	0.19	86
d_9	1	0.93	0.37	−0.07	90	0.94	0.38	−0.06	71
d_{10}	1	0.91	0.40	−0.09	96	0.96	0.39	−0.04	81
d_{11}	2	1.76	0.52	−0.24	85	1.55	0.48	−0.45	69
d_{12}	2	1.78	0.51	−0.22	91	1.45	0.47	−0.55	57
d_{13}	3	2.75	0.71	−0.25	84	1.96	0.53	−1.04	47
d_{14}	3	2.79	0.67	−0.21	86	1.94	0.51	−1.06	41
d_{15}	4	3.95	1.20	−0.05	84	2.15	0.57	−1.85	28
d_{16}	4	3.80	1.09	−0.20	89	2.13	0.56	−1.87	22
ν	1	0.83	0.24	−0.17	85	1.06	0.31	0.06	93
ω	1	0.82	0.19	−0.18	81	0.82	0.20	−0.18	77
σ_b^2	0.10					0.13	0.03	0.03	98

σ_d^2	3.00					2.68	0.44	−0.32	52

Table notes: Par. is the model parameter, Pop. is the population value, Av. PM is the average posterior mean over 100 replications, Av. PSD is the average posterior standard deviation over 100 replications, % Cov. is the obtained 90% coverage rate.

Table A2

Parameter Recovery for Simulated Data, Bimodal λ_i, N=500

		IRSDT				IRSDTr			
Par.	Pop.	Av. PM	Av. PSD	Bias	% Cov.	Av. PM	Av. PSD	Bias	% Cov.
b_1	0	0.04	0.16	0.04	93	−0.06	0.20	−0.06	83
b_2	0	−0.05	0.16	−0.05	92	0.01	0.21	0.01	89
b_3	0	0.05	0.16	0.05	95	−0.03	0.21	−0.03	87
b_4	0	−0.09	0.17	−0.09	86	−0.05	0.21	−0.05	82
b_5	0	0.07	0.17	0.07	91	0.01	0.20	0.01	83
b_6	0	−0.05	0.17	−0.05	97	0.01	0.20	0.01	80
b_7	0	−0.01	0.16	−0.01	93	0.00	0.18	0.00	90
b_8	0	−0.01	0.16	−0.01	92	−0.04	0.18	−0.04	88
b_9	1	1.03	0.20	0.03	96	1.00	0.19	0.00	94
b_{10}	−1	−1.08	0.20	−0.08	93	−0.99	0.19	0.01	85
b_{11}	1.5	1.65	0.31	0.15	88	1.48	0.26	−0.02	91
b_{12}	−1.5	−1.58	0.28	−0.08	94	−1.50	0.25	0.00	93
b_{13}	2	2.36	0.45	0.36	90	2.07	0.33	0.07	81
b_{14}	−2	−2.38	0.46	−0.38	91	−2.09	0.35	−0.09	80
b_{15}	0.5	0.62	0.19	0.12	89	0.44	0.23	−0.06	77
b_{16}	−0.5	−0.58	0.19	−0.08	95	−0.49	0.24	0.01	80
d_1	4	4.80	1.09	0.80	88	3.00	0.56	−1.00	51
d_2	4	4.52	1.06	0.52	84	3.13	0.56	−0.87	63
d_3	3	3.79	0.92	0.79	88	2.55	0.54	−0.45	68
d_4	3	3.99	0.97	0.99	82	2.70	0.52	−0.30	69
d_5	2	2.30	0.49	0.30	90	1.76	0.46	−0.24	70
d_6	2	2.17	0.44	0.17	96	1.73	0.45	−0.27	72
d_7	1	1.00	0.29	0.00	94	0.84	0.35	−0.16	82
d_8	1	1.03	0.29	0.03	92	0.89	0.34	−0.11	80
d_9	1	1.04	0.29	0.04	92	0.72	0.30	−0.28	80
d_{10}	1	1.09	0.30	0.09	89	0.74	0.30	−0.26	71
d_{11}	2	2.18	0.41	0.18	90	1.64	0.41	−0.36	72
d_{12}	2	2.08	0.37	0.08	96	1.60	0.39	−0.40	72
d_{13}	3	3.41	0.54	0.41	88	2.67	0.47	−0.33	74
d_{14}	3	3.46	0.56	0.46	89	2.65	0.48	−0.35	71
d_{15}	4	5.37	1.13	1.37	81	3.09	0.56	−0.91	57
d_{16}	4	5.17	1.09	1.17	84	3.18	0.57	−0.82	55
ν	0.3	0.43	0.07	0.13	37	0.46	0.08	0.16	24
ω	0.3	0.41	0.06	0.11	27	0.36	0.05	0.06	72

σ_b^2	0.1	0.13	0.03	0.03	94
σ_d^2	3.0	3.12	0.47	0.12	83

Table notes: Par. is the model parameter, Pop. is the population value, Av. PM is the average posterior mean over 100 replications, Av. PSD is the average posterior standard deviation over 100 replications, % Cov. is the obtained 90% coverage rate.