Supplementary materials for *Using Auxiliary Item Information in the Item Parameter Estimation of a Graded Response Model for a Small to Medium Sample Size: Empirical versus Hierarchical Bayes Estimation*

Matthew Naveiras and Sun-Joo Cho
Vanderbilt University

Appendix A

Probability Density Functions of Different Prior Distributions

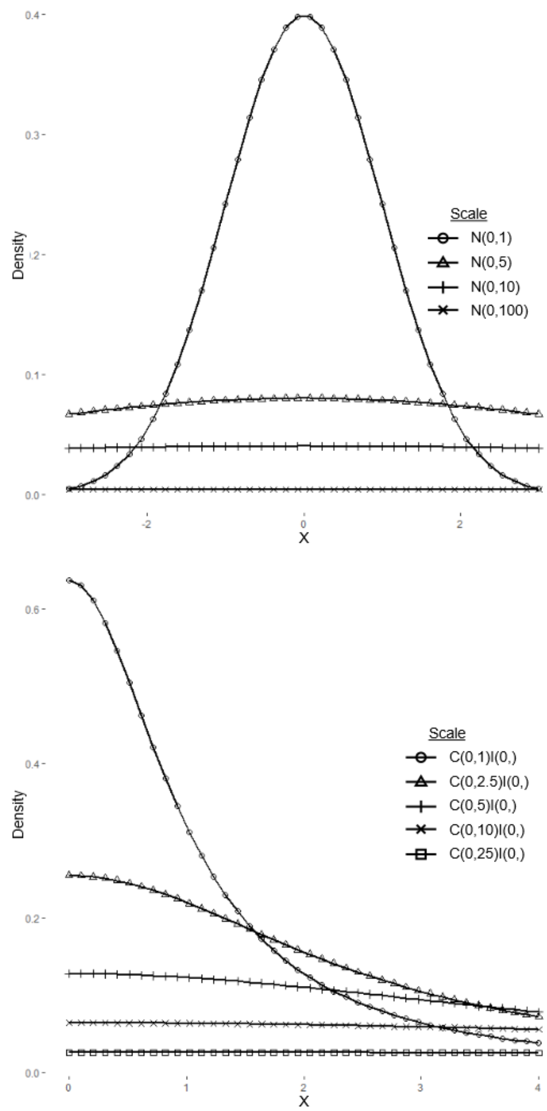

Figure A1: Probability density functions of different normal (top) and half-Cauchy (bottom) distribution scales

Appendix B

Empirical Study

Illustration

In this section, we illustrate the empirical and hierarchical Bayes methods described by applying them to an empirical data set. `R` functions to implement the methods used below are available on GitHub (https://github.com/naveirmd/Auxiliary_Item_Information_GRM).

Data Description

The data analyzed using the methods described were collected from the Vanderbilt Fatigue Scale for Adults (VFS-A), which was designed to measure listening-related fatigue. Preliminary research led to the identification of four domains of listening-related fatigue: cognitive, emotional, physical, and social (Davis, Schlundt, Camarata, Bess, & Hornsby, 2020). Using Mplus Version 8.3 (Muthén & Muthén, 1998-2017), exploratory factor analyses were conducted using polychoric correlations (specifically, limited information robust weighted least square estimation with Oblimin rotation and Oblique type) to extract 1, 2, 3, and 4 factors[1] to explore the number and structure of the factors of the VFS-AHL. In Table B1 these four models were compared with standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI). Based on empirically-supported guidelines a model is considered to fit well if $SRMR < .08$, $RMSEA < .06$, $CFI > .95$, and $TLI > .95$ (Hu & Bentler, 1999; Yu, 2002). The unidimensional model was considered a well-fitting model according to the SRMR, CFI, and TLI. In addition, parallel analysis with polychoric correlations (Cho, Li, & Bandalos, 2009) supported the extraction of one dimension. Based on the exploratory factor analyses and the parallel analysis, we considered listening-related fatigue to be a unidimensional construct. Based on these exploratory factor analyses, we considered listening-related fatigue to be a unidimensional construct.

[1] Quartimin, Geomin, and Target rotation methods resulted in similar patterns of factor loadings across all methods.

Table B1: Fit Indices from Exploratory Factor Analyses

Fix Indices	1-Factor	2-Factor	3-Factor	4-Factor
SRMR	0.038	0.028	0.022	0.018
RMSEA	0.082[0.079,0.085]*	0.072[0.069,0.075]*	0.062[0.059,0.064]*	0.054[0.051,0.057]*
CFI	0.986	0.990	0.993	0.995
TLI	0.985	0.989	0.992	0.994

Note. * 90% confidence interval

The research version of the VFS-A was analyzed, having 10 five-point Likert-scale items for each of the four domains of listening-related fatigue, for a total of 40 items. A total of 273 participants completed all 40 items. Of these 273 participants, 150 participants were randomly sampled to illustrate the empirical and hierarchical Bayes methods for a small sample size.

Analysis

The VFS-A has a mutually-exclusive binary Q-matrix having four domains: cognitive, emotional, physical, and social. The four domains were treated as item covariates for analysis.

For dummy variable coding, the social domain was chosen (arbitrarily) as the reference category. The regression structure for item parameters was structured as follows:

$$\alpha_i = \gamma_{\alpha 0} + \gamma_{\alpha 1} x_{i1} + \gamma_{\alpha 2} x_{i2} + \gamma_{\alpha 3} x_{i3} + \epsilon_{\alpha i} \tag{1}$$

and

$$\beta_{i,k} = \gamma_{\beta 0k} + \gamma_{\beta 1k} x_{i1} + \gamma_{\beta 2k} x_{i2} + \gamma_{\beta 3k} x_{i3} + \epsilon_{\beta ik}, \tag{2}$$

where $x_{i1} = 1$ for cognitive items, $x_{i2} = 1$ for emotional items, and $x_{i3} = 1$ for physical items. Estimates for the regression coefficients ($\gamma_{\alpha 0}$, $\gamma_{\alpha d}$, $\gamma_{\beta 0k}$, and $\gamma_{\beta dk}$) were obtained from the `lm` function in `R`, using maximum likelihood estimates obtained from `mirt`.

The regression structures for hierarchical Bayesian priors were structured as follows:

$$\alpha_i \sim N(\gamma_{\alpha 0} + \gamma_{\alpha 1} x_{i1} + \gamma_{\alpha 2} x_{i2} + \gamma_{\alpha 3} x_{i3}, \sigma^2_{\alpha}) \tag{3}$$

and

$$\beta_{i,k} \sim N(\gamma_{\beta 0k} + \gamma_{\beta 1k} x_{i1} + \gamma_{\beta 2k} x_{i2} + \gamma_{\beta 3k} x_{i3}, \sigma^2_{\beta k}). \tag{4}$$

For regression coefficients ($\gamma_{\alpha 0}$, $\gamma_{\alpha d}$, $\gamma_{\beta 0k}$, and $\gamma_{\beta dk}$), a non-informative hyper-prior distribution of $N(0, 10^2)$ was chosen. A weakly-informative hyper-prior distribution of $Cauchy(0, 10)I(0,)$ was imposed on the standard deviations of residuals (σ_α and $\sigma_{\beta k}$).

The default arguments for `rStan` of 4 chains, $2,000$ iterations, 1,000 warmup (i.e., burn-in) iterations per chain, and thinning = 1 were used for analysis. Convergence amongst the 4 chains was evaluated using the Gelman-Rubin statistic (Gelman & Rubin, 1992). Note that these arguments were sufficient to achieve sufficient convergence in both sample sizes, having Gelman-Rubin statistics in the range of 0.95 to 1.05 for all parameters. Obtaining results for sample sizes of 150 in `R` required approximately 1.2 hours on a computer with a 2.30GHz processor and 8.00gb of RAM.

Results

The results obtained for $J = 150$ are presented in this section. Comparisons of regression coefficient estimates for both empirical Bayes and hierarchical Bayes methods are illustrated in Table B2. Median hierarchical Bayes estimates were used for calculating hierarchical shrinkage and standardized differences, as well as for comparing hierarchical and empirical Bayes methods. Results were highly comparable between empirical and hierarchical Bayes methods: $r(18) = .995$, p-value $< .01$ for item regression parameters, and $r(3) = .975$, p-value $< .01$ for the standard deviations of residuals ($\widehat{\phi}_\alpha$ and $\widehat{\phi}_{\beta k}$). However, the standard deviation of residuals were generally larger for the empirical Bayes method than for the hierarchical Bayes method.

Table B3 reports the results for empirical and hierarchical Bayes estimates of α_i, and Table B4 reports the results for empirical and hierarchical Bayes estimates of β_{i4} for illustration. The standard deviations of empirical Bayes estimates were lower than the standard errors of maximum likelihood estimates, because of the added information from item covariates. However, even for a small sample size of J=150, maximum likelihood estimates had significantly lower standard errors than regression estimates (as seen in Table B3 by comparing $\widehat{\tau}_{\alpha i}$ and $\bar{\phi}_\alpha$), because the information provided by the data far outweighed the information provided by the item covariates. This is

Table B2: Regression Coefficients for Empirical Bayes vs. Hierarchical Bayes, J=150

	Empirical Bayes		Hierarchical Bayes				
	EST	SE	Mean	Median	SD	0.025*	0.975*
Discrimination							
$\gamma_{\alpha 0}$	2.095	0.251	2.143	2.135	0.281	1.620	2.719
$\gamma_{\alpha 1}$	0.673	0.365	0.509	0.509	0.361	-0.229	1.208
$\gamma_{\alpha 2}$	0.979	0.355	0.999	0.993	0.368	0.285	1.736
$\gamma_{\alpha 3}$	0.458	0.355	0.319	0.313	0.355	-0.380	1.019
ϕ_{α}	0.794	0.147	0.741	0.732	0.112	0.546	0.986
Threshold 1							
$\gamma_{\beta 01}$	-4.149	0.593	-4.001	-3.989	0.558	-5.095	-2.912
$\gamma_{\beta 11}$	-2.116	0.861	-2.168	-2.173	0.782	-3.742	-0.632
$\gamma_{\beta 21}$	-1.356	0.838	-1.303	-1.317	0.767	-2.777	0.200
$\gamma_{\beta 31}$	0.142	0.838	0.500	0.510	0.767	-1.023	2.046
$\phi_{\beta 1}$	1.874	0.820	1.619	1.602	0.236	1.216	2.135
Threshold 2							
$\gamma_{\beta 02}$	-2.179	0.423	-1.997	-2.000	0.423	-2.795	-1.175
$\gamma_{\beta 12}$	-1.874	0.615	-1.896	-1.889	0.571	-3.050	-0.775
$\gamma_{\beta 22}$	-1.041	0.598	-0.973	-0.966	0.552	-2.090	0.137
$\gamma_{\beta 32}$	0.548	0.598	0.714	0.727	0.551	-0.395	1.767
$\phi_{\beta 2}$	1.338	0.418	1.202	1.187	0.162	0.928	1.558
Threshold 3							
$\gamma_{\beta 03}$	-0.249	0.353	-0.071	-0.073	0.403	-0.865	0.717
$\gamma_{\beta 13}$	-1.238	0.513	-1.409	-1.410	0.527	-2.426	-0.379
$\gamma_{\beta 23}$	-0.522	0.499	-0.461	-0.466	0.527	-1.486	0.573
$\gamma_{\beta 33}$	0.722	0.499	0.772	0.763	0.523	-0.253	1.775
$\phi_{\beta 3}$	1.116	0.290	1.110	1.098	0.144	0.867	1.435
Threshold 4							
$\gamma_{\beta 04}$	1.653	0.314	1.811	1.815	0.382	1.070	2.555
$\gamma_{\beta 14}$	-0.381	0.456	-0.627	-0.631	0.492	-1.598	0.345
$\gamma_{\beta 24}$	0.081	0.444	0.158	0.162	0.497	-0.825	1.109
$\gamma_{\beta 34}$	1.009	0.444	0.970	0.967	0.502	-0.022	1.932
$\phi_{\beta 4}$	0.993	0.230	1.067	1.051	0.142	0.835	1.382

Note. * Percentiles of posterior distribution

further seen in the calculated values for shrinkage. The average shrinkage for items 2-40 in Table B3 (note that shrinkage was not calculated for item 1) was .159, meaning that on average item covariates contributed 15.9% of the information used in estimating α_i, with the data providing the remaining 84.1% of the information. Similar results were obtained for item threshold parameters. Item threshold parameters (β_{i1}, β_{i2}, β_{i3}, and β_{i4}) had average shrinkages across items of 11.2%, 10.2%, 9.1%, and 12.7% (respectively). The item parameter estimates obtained using empirical Bayes and hierarchical Bayes (as well as their respective SDs or SEs) were highly similar for all item parameters at each sample size.

Table B3: Comparison of Empirical and Hierarchical Bayesian Estimates for α_i with J=150

		Empirical							Hierarchical	
		Step 1		Step 2		Step 3				
Item	Domain	$\widehat{\alpha}_i$	$\widehat{\tau}_{\alpha i}$	$\bar{\alpha}_i$	ϕ_α	$\tilde{\alpha}_i$	$\tilde{\sigma}_{\alpha i}$	Shrinkage	Posterior Median	SD
1	C	-5.361	0.802	NA	NA	NA	NA	NA	-7.180	1.263
2	C	-8.430	1.250	-6.265	1.874	-7.763	1.040	0.308	-7.628	0.925
3	C	-8.382	1.068	-6.265	1.874	-7.863	0.928	0.245	-7.267	0.760
4	C	-6.893	1.121	-6.265	1.874	-6.728	0.962	0.263	-6.701	0.936
5	C	-7.202	0.921	-6.265	1.874	-7.020	0.826	0.194	-6.705	0.726
6	C	-5.978	0.723	-6.265	1.874	-6.015	0.674	0.129	-6.056	0.660
7	C	-4.747	0.594	-6.265	1.874	-4.886	0.567	0.091	-4.573	0.532
8	C	-5.924	0.758	-6.265	1.874	-5.972	0.702	0.141	-6.094	0.702
9	C	-4.927	0.596	-6.265	1.874	-5.050	0.568	0.092	-4.984	0.531
10	C	-3.901	0.458	-6.265	1.874	-4.034	0.445	0.056	-4.156	0.478
11	E	-4.324	0.556	-5.505	1.874	-4.419	0.533	0.081	-4.600	0.575
12	E	-5.863	0.704	-5.505	1.874	-5.819	0.659	0.124	-5.657	0.622
13	E	-4.778	0.565	-5.505	1.874	-4.838	0.541	0.083	-4.920	0.554
14	E	-5.523	0.674	-5.505	1.874	-5.521	0.635	0.115	-5.271	0.575
15	E	-6.735	0.820	-5.505	1.874	-6.537	0.751	0.161	-6.273	0.654
16	E	-3.166	0.456	-5.505	1.874	-3.297	0.443	0.056	-3.139	0.424
17	E	-6.319	0.757	-5.505	1.874	-6.205	0.702	0.140	-5.966	0.623
18	E	-7.211	0.880	-5.505	1.874	-6.903	0.797	0.181	-6.536	0.695
19	E	-5.848	0.755	-5.505	1.874	-5.800	0.700	0.140	-5.380	0.599
20	E	-5.284	0.632	-5.505	1.874	-5.307	0.599	0.102	-5.156	0.553
21	P	-8.197	1.044	-4.007	1.874	-7.205	0.912	0.237	-6.644	0.738
22	P	-7.240	0.937	-4.007	1.874	-6.594	0.838	0.200	-6.120	0.688
23	P	-2.773	0.392	-4.007	1.874	-2.825	0.383	0.042	-2.695	0.369
24	P	-7.667	0.966	-4.007	1.874	-6.898	0.859	0.210	-6.150	0.657
25	P	-5.524	0.692	-4.007	1.874	-5.342	0.649	0.120	-4.766	0.539
26	P	-2.033	0.319	-4.007	1.874	-2.089	0.314	0.028	-1.962	0.297
27	P	-1.330	0.228	-4.007	1.874	-1.369	0.226	0.015	-1.357	0.230
28	P	-2.736	0.389	-4.007	1.874	-2.789	0.381	0.041	-2.640	0.358
29	P	-1.115	0.237	-4.007	1.874	-1.160	0.235	0.016	-1.124	0.230
30	P	-1.453	0.287	-4.007	1.874	-1.511	0.284	0.023	-1.413	0.270
31	S	-6.171	0.841	-4.149	1.874	-5.832	0.767	0.168	-5.709	0.704
32	S	-3.524	0.413	-4.149	1.874	-3.553	0.404	0.046	-3.535	0.391
33	S	-6.904	0.849	-4.149	1.874	-6.435	0.773	0.170	-6.057	0.662
34	S	-4.947	0.595	-4.149	1.874	-4.874	0.567	0.091	-4.620	0.527
35	S	-3.422	0.406	-4.149	1.874	-3.454	0.396	0.045	-3.430	0.383
36	S	-3.266	0.401	-4.149	1.874	-3.304	0.392	0.044	-3.242	0.387
37	S	-3.367	0.448	-4.149	1.874	-3.409	0.436	0.054	-3.173	0.401
38	S	-3.218	0.389	-4.149	1.874	-3.257	0.381	0.041	-3.239	0.376
39	S	-3.026	0.390	-4.149	1.874	-3.073	0.382	0.042	-2.964	0.367
40	S	-3.647	0.430	-4.149	1.874	-3.672	0.419	0.050	-3.718	0.411

Note. Maximum likelihood estimation was unable to estimate the fourth threshold of item 1 ($\beta_{1,4}$) because no responses were obtained in the fifth category of item 1. Because of this, item 1 was omitted during Step 2 when obtaining regression estimates, resulting in the missing values for item 1 (noted as NA). Because regression estimates were required for the values calculated in Step 3, these values are also missing for item 1. This is one instance where hierarchical Bayes estimation has an advantage over empirical Bayes estimation. Because the hyper-prior distribution for $\beta_{1,4}$ in hierarchical Bayesian estimation provides information outside of the data, $\beta_{1,4}$ is still capable of being estimated.

Appendix C

LLTM Literature Review

Papers published in these six journals were reviewed to report how item covariate structures were used for item response models in common practice: *Acta Psychologica*, *Applied Psychological Measurement (APM)*, *Educational and Psychological Measurement (EPM)*, *Journal of Educational Measurement (JEM)*, *Multivariate Behavioral Research (MBR)*, and *Psychometrika (PMET).* Papers were searched using the keywords "item response theory linear logistic test model."

Q-matrices generally took on one of four common patterns. First, 36% (10) of the reviewed papers used a *mutually-exclusive binary Q-matrix* item covariate structure. Items constructed in this way were assigned a value of 1 for at most one of the item covariates, and a value of 0 for all other item covariates. Second, 25% (7) of the reviewed papers used a *non-mutually-exclusive binary Q-matrix* item covariate structure. Items constructed in this way were assigned a value of 1 for any number of item covariates, and a value of 0 for all other item covariates. Third, 11% (3) of the reviewed papers used a *non-mutually-exclusive non-binary Q-matrix* item covariate structure. Items constructed in this way were assigned a value to each item covariate indicative of how many occurrences of that item trait were present in the item. Fourth, 29% (8) of the reviewed papers used a *Q-matrix by factor* item covariate structure. Items constructed in this way had one of the three previously defined item covariate structures for each of two or more item factors, although the most common pattern of this structure was a mutually-exclusive binary Q-matrix for each factor.

Table B4: Comparison of Empirical and Hierarchical Bayesian Estimates for $\beta_{i,4}$ with J=150

		Empirical							Hierarchical	
		Step 1		Step 2		Step 3				
Item	Domain	$\widehat{\beta}_{i,4}$	$\widehat{\tau}_{\beta i,4}$	$\bar{\beta}_{i,4}$	$\bar{\phi}_{\beta,4}$	$\tilde{\beta}_{i,4}$	$\tilde{\sigma}_{\beta i,4}$	Shrinkage	Posterior Median	SD
1	C	NA	NA	NA	NA	NA	NA	NA	-1.184	0.247
2	C	0.712	0.352	1.272	0.993	0.774	0.332	0.111	0.927	0.312
3	C	2.128	0.499	1.272	0.993	1.956	0.446	0.202	2.150	0.400
4	C	-0.306	0.285	1.272	0.993	-0.186	0.274	0.076	-0.067	0.257
5	C	0.611	0.368	1.272	0.993	0.691	0.345	0.121	0.829	0.320
6	C	1.347	0.339	1.272	0.993	1.339	0.320	0.104	1.548	0.323
7	C	3.144	0.483	1.272	0.993	2.785	0.435	0.191	3.154	0.421
8	C	1.366	0.312	1.272	0.993	1.358	0.298	0.090	1.555	0.301
9	C	1.079	0.341	1.272	0.993	1.100	0.323	0.105	1.306	0.326
10	C	1.364	0.280	1.272	0.993	1.357	0.270	0.074	1.522	0.283
11	E	0.417	0.215	1.735	0.993	0.476	0.210	0.045	0.586	0.221
12	E	2.078	0.392	1.735	0.993	2.032	0.365	0.135	2.287	0.361
13	E	1.659	0.315	1.735	0.993	1.666	0.300	0.091	1.913	0.316
14	E	1.396	0.409	1.735	0.993	1.445	0.378	0.145	1.682	0.371
15	E	2.127	0.446	1.735	0.993	2.061	0.407	0.168	2.307	0.397
16	E	2.875	0.429	1.735	0.993	2.696	0.394	0.157	3.058	0.400
17	E	1.822	0.421	1.735	0.993	1.809	0.387	0.152	2.036	0.382
18	E	2.297	0.478	1.735	0.993	2.191	0.431	0.188	2.432	0.410
19	E	1.093	0.461	1.735	0.993	1.207	0.418	0.177	1.405	0.401
20	E	1.581	0.391	1.735	0.993	1.602	0.364	0.134	1.862	0.370
21	P	2.213	0.451	2.662	0.993	2.290	0.411	0.171	2.268	0.378
22	P	-0.226	0.358	2.662	0.993	0.107	0.337	0.115	0.210	0.284
23	P	2.399	0.365	2.662	0.993	2.430	0.342	0.119	2.647	0.347
24	P	3.550	0.562	2.662	0.993	3.335	0.489	0.242	3.315	0.434
25	P	3.128	0.519	2.662	0.993	3.028	0.460	0.214	3.138	0.433
26	P	3.986	0.467	2.662	0.993	3.747	0.422	0.181	3.998	0.418
27	P	2.201	0.279	2.662	0.993	2.235	0.269	0.073	2.375	0.283
28	P	2.903	0.395	2.662	0.993	2.870	0.367	0.137	3.105	0.378
29	P	2.873	0.345	2.662	0.993	2.850	0.326	0.108	2.996	0.328
30	P	3.590	0.427	2.662	0.993	3.446	0.392	0.156	3.654	0.381
31	S	0.700	0.281	1.653	0.993	0.771	0.270	0.074	0.909	0.265
32	S	1.423	0.280	1.653	0.993	1.440	0.269	0.074	1.628	0.271
33	S	1.047	0.369	1.653	0.993	1.121	0.346	0.122	1.252	0.313
34	S	-0.269	0.282	1.653	0.993	-0.126	0.271	0.075	0.004	0.253
35	S	2.024	0.300	1.653	0.993	1.993	0.287	0.084	2.181	0.295
36	S	2.067	0.319	1.653	0.993	2.028	0.304	0.094	2.251	0.304
37	S	3.449	0.446	1.653	0.993	3.147	0.407	0.168	3.411	0.405
38	S	2.083	0.301	1.653	0.993	2.047	0.288	0.084	2.234	0.298
39	S	2.801	0.373	1.653	0.993	2.659	0.349	0.123	2.859	0.342
40	S	1.206	0.257	1.653	0.993	1.234	0.248	0.063	1.396	0.251

Note. Maximum likelihood estimation was unable to estimate the fourth threshold of item 1 ($\beta_{1,4}$) because no responses were obtained in the fifth category of item 1. Because of this, item 1 was omitted during Step 2 when obtaining regression estimates, resulting in the missing values for item 1 (noted as NA).

Table C1: *LLTM Literature Review*

Reference	Number of Items	Number of Item Covariates (per factor)	Item Covariate Structure
Baker (1993)	21	8	Non-mutually exclusive binary Q-matrix
Bechger, Verstralen, & Verhelst (2002)	5	2	Non-mutually exclusive non-binary Q-matrix
Beretvas & Williams (2004)	17	2	Mutually exclusive binary Q-matrix
Bolt, Cohen, & Wollack (2002)	26	2	Mutually exclusive binary Q-matrix
Chalmers (2015)	15	3	Mutually exclusive binary Q-matrix
Choi & Wilson (2015)	24	2x2x3=12**	Q-matrix by factor
De Boeck (2008)	24	2x2x3=12**	Q-matrix by factor
Embretson (2015)	70	4	Non-mutually exclusive binary Q-matrix
Fischer (1973)	29	8	Non-mutually exclusive binary Q-matrix
Freund, Hofer, & Holling (2008)	25	5	Non-mutually exclusive binary Q-matrix
Gorin (2005)	29	5	Mutually exclusive binary Q-matrix
Hartig, Frey, Nold, & Klieme (2012)	46	2x2=4**	Q-matrix by factor
Hoffman, Yang, Bovaird, & Embretson (2006)	64	3	Non-mutually exclusive non-binary Q-matrix
Hohensinn & Kubinger (2011)	18	3	Mutually exclusive binary Q-matrix
Hornke & Habon (1986)	24	8x3=24**	Q-matrix by factor
Ip, Magee, Youssef, & Chen (2019)	31	6	Non-mutually exclusive binary Q-matrix
Ip, Smits, & De Boeck (2009)	8	2x2=4**	Q-matrix by factor
Kim (2018)	13	3x3x3=27**	Q-matrix by factor
Kubinger (2008)	29	8	Non-mutually exclusive binary Q-matrix
Medina-Diaz (1993)	29	8	Non-mutually exclusive binary Q-matrix
Mislevy (1988)	20	6	Mutually exclusive binary Q-matrix
Mitchell (1983)	334	10	Mutually exclusive binary Q-matrix
Poinstingl (2009)	25	8***	Q-matrix by factor
Rakkapao, Prasitpong, & Arayathanitkul (2016)	20	10	Mutually exclusive binary Q-matrix
Rost & Cartensen (2002)	77	11x7=77**	Q-matrix by factor
Sheehan & Mislevy (1990)	93	3	Mutually exclusive binary Q-matrix
Shermis & Chang (1997)	45/90/90*	4	Mutually exclusive binary Q-matrix
Whitely & Schneider (1981)	30	8	Non-mutually exclusive non-binary Q-matrix

Note. * Three forms of the same test were used in this study. Form A had 45 items, and Forms B/C each had 90 items.
** The number of item groups is equal to the product of the number of levels per factor. For example, three factors with two levels for each factor results in $2 \times 2 \times 2 = 8$ item groups.
*** The number of factors in this study was 8. Three factors had mutually exclusive binary Q-matrices with 4, 3, and 5 levels. The remaining five factors counted the occurrences of different item attributes.

Table C2: R^2 *of Simulated Item Parameters for Each Level of RSD*

		ME Q-matrix		NME Q-matrix	
		24 items	48 items	24 items	48 items
RSD	.10	.81	.83	.92	.84
	.30	.48	.44	.57	.67
	.50	.17	.26	.40	.45

Appendix D

Hypotheses of Simulation Results

Research question 1a: Accuracy of item parameter estimates. Because the use of group means results in shrinkage, which increases the accuracy of item parameter estimates (see p. 13), we expect both empirical Bayes and hierarchical Bayes to have lower RPB and lower RMSE than MMLE (which does not use group means at all, and therefore has no shrinkage). Because we expect MMLE to have high RMSE at small and medium sample sizes, we also expect empirical Bayes (which uses maximum likelihood estimates in Step 2) to have higher RPB and higher RMSE than hierarchical Bayes with item covariates (which does not use maximum likelihood estimates). Therefore, we expect the following relations regarding RPB and RMSE: empirical Bayes < MMLE, hierarchical Bayes with item covariates < MMLE, and hierarchical Bayes with item covariates < empirical Bayes. For medium to large sample sizes, RPB and RMSE are comparable.

Research question 1b: Acceptability of hierarchical Bayes. As discussed in the introduction, MMLE is expected to have difficulty with achieving convergence in smaller sample sizes, making estimation of item parameters impossible. In such conditions we expect a hierarchical Bayes method with item covariates to estimate item parameters and posterior SD with an acceptable degree of accuracy, having $RPB < 10\%$ and $SDB < 10\%$ for all item parameter types.

Research question 2: Added accuracy of item covariates. To examine the added value of item covariates in a hierarchical Bayes method for all conditions, we compare the accuracy of a hierarchical Bayes method both with and without item covariates. As discussed previously (see p. 13), because the use of multiple item covariates results in group shrinkage rather than total shrinkage, it is expected that hierarchical Bayes with item covariates will have lower RPB than hierarchical Bayes without item covariates. Therefore, regarding RPB we expect hierarchical Bayes with item covariates < hierarchical Bayes without item covariates.[2]

[2]Although we expect lower RMSE when item covariates are used (in empirical Bayes and hierarchical Bayes methods) compared to when item covariates are not used (in MMLE), we have no such expectations for RMSE regarding the change in RMSE by using multiple item covariates (in hierarchical Bayes with item covariates) as

Research question 3: Accuracy of posterior SD estimates. Because empirical Bayes ignores the uncertainty of item parameter estimates in Step 2, it is expected that empirical Bayes will result in large SDB. Alternatively, because hierarchical Bayes implements this uncertainty by using a one-step approach, it is expected that hierarchical Bayes will more accurately estimate its posterior SD than empirical Bayes, resulting in a smaller SDB. Therefore, for SDB we expect hierarchical Bayes with item covariates < empirical Bayes.

In addition to these hypotheses, certain patterns are expected across estimation methods regarding the four simulation factors:

Number of persons. It is expected that an increase in the number of persons will result in decreases in RPB and RMSE. As the number of persons approaches a large sample size of 2000, differences in the RMSE among these methods are expected to decrease as the accuracy of MMLE (which most prominently suffers in small sample sizes) increases. It is also expected than an increase in the number of persons will result in a decrease in SDB for both empirical Bayes and hierarchical Bayes.

Number of items. It is expected that an increase in the number of items will result in a decrease in the RPB as prior means are based on a larger number of items, therefore decreasing the shrinkage for individual items. Alternatively, it is expected that an increase in the number of items will result in higher RMSE, as there are more item parameters to estimate. For a fixed number of persons, it is expected that an increase in the number of items will result in higher SDB.

RSD. It is expected that an increase in RSD will result in a decrease in the RPB and an increase in the RMSE, because less shrinkage is expected with a larger RSD in empirical Bayes and hierarchical Bayes. A larger RSD is not expected to affect SDB.

Item covariate structure. It is expected that RPB and RMSE are larger for ME Q-matrix

---

opposed to using a single item covariates (in hierarchical Bayes without item covariates). As a result, we do not have any hypotheses regarding differences in RMSE between hierarchical Bayes with item covariates and hierarchical Bayes without item covariates.

conditions than for NME Q-matrix conditions because there is a smaller number of items having the same item covariate in the ME Q-matrix (sparse matrix) than in the NME Q-matrix (dense matrix).

Appendix E

Results of Simulation Results

Table E1: *RPB of MMLE, Empirical Bayes, & Hierarchical Bayes by Item Parameter Type for Research Question 1a*

				MMLE					Empirical Bayes					Hierarchical Bayes				
# Items	RSD	# Persons	Q-matrix	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}
24	0.1	150	ME	1.709	1.517	1.107	2.739	2.557	-34.020	-39.749	-63.628	-13.104	-19.702	4.154	0.911	0.625	1.993	1.713
24	0.1	200	ME	1.340	1.872	1.764	1.034	1.417	-33.123	-35.476	-59.218	-13.033	-18.852	3.309	1.303	1.411	0.574	0.806
24	0.1	250	ME	0.795	1.154	0.823	1.417	1.385	-32.417	-34.127	-58.262	-11.881	-17.625	2.430	0.688	0.618	0.926	0.805
24	0.1	300	ME	0.583	1.222	1.059	0.805	1.063	-31.597	-32.036	-54.943	-11.426	-16.432	2.031	0.887	1.139	0.255	0.487
24	0.1	500	ME	0.375	0.944	0.947	0.709	0.829	-27.868	-24.751	-46.576	-8.977	-12.675	1.302	0.637	0.874	0.329	0.445
24	0.1	2000	ME	0.166	0.513	0.575	0.346	0.390	-16.452	-8.954	-18.941	-5.733	-5.747	0.309	0.372	0.492	0.133	0.218
24	0.3	300	ME	0.547	5.030	3.866	7.129	5.872	-9.344	-0.777	-7.945	-9.042	1.369	2.904	3.810	-0.799	-6.807	4.590
24	0.3	500	ME	0.352	5.033	4.538	6.476	5.333	-6.248	1.371	-3.008	-4.473	2.441	1.712	4.309	1.775	-1.814	4.441
24	0.3	2000	ME	0.001	4.568	4.532	5.236	4.712	-2.047	3.614	2.867	3.428	3.712	0.374	4.426	3.907	3.192	4.446
24	0.5	2000	ME	-0.323	-1.399	-0.324	-1.580	-1.430	-1.343	-2.098	-1.679	-2.165	-2.065	0.342	-1.532	-0.183	-1.383	-1.503
48	0.1	150	ME	1.271	5.817	5.464	5.927	6.117	-31.468	-32.854	-50.321	-10.012	-16.339	3.295	4.380	4.526	5.878	4.789
48	0.1	200	ME	1.081	5.246	4.916	5.646	5.751	-30.479	-30.136	-47.551	-9.209	-15.130	2.792	4.179	4.430	5.470	4.653
48	0.1	250	ME	0.572	5.338	5.443	4.480	5.060	-29.961	-27.633	-44.373	-9.043	-13.884	2.063	4.442	5.062	4.385	4.132
48	0.1	300	ME	0.679	5.044	4.805	4.890	5.045	-28.804	-25.362	-41.550	-8.149	-12.764	1.998	4.236	4.421	4.850	4.290
48	0.1	500	ME	0.358	4.588	4.542	4.721	4.753	-25.295	-19.178	-33.500	-5.830	-9.166	1.271	3.994	4.222	4.661	4.282
48	0.1	2000	ME	0.108	4.345	4.409	4.179	4.253	-13.745	-4.189	-9.499	-3.396	-2.203	0.428	4.115	4.158	4.179	4.152
48	0.3	500	ME	0.358	-2.312	-2.609	-1.744	-1.855	-7.131	-7.672	-10.690	-4.203	-5.382	2.177	-3.043	-2.447	-2.542	-2.809
48	0.3	2000	ME	-0.345	-2.543	-2.563	-2.520	-2.511	-3.087	-4.035	-4.486	-3.712	-3.854	0.297	-2.744	-2.573	-2.735	-2.750
24	0.1	150	NME	1.007	2.021	1.559	2.071	2.342	-22.153	-21.125	-30.265	-16.268	-15.090	2.409	2.209	1.232	0.001	1.521
24	0.1	200	NME	1.305	1.772	1.570	1.366	1.829	-20.881	-19.400	-28.541	-16.136	-13.803	2.178	2.204	1.806	-1.024	0.882
24	0.1	250	NME	1.129	1.441	1.162	1.296	1.449	-20.422	-18.427	-27.481	-15.195	-12.917	1.814	2.137	2.065	-1.603	0.285
24	0.1	300	NME	0.863	1.324	1.360	0.968	1.231	-19.869	-16.998	-25.882	-14.636	-12.160	1.347	2.091	2.503	-1.906	0.100
24	0.1	500	NME	0.513	1.002	0.899	0.706	0.872	-17.585	-12.939	-21.942	-12.404	-9.223	0.802	1.886	2.456	-2.155	-0.268
24	0.1	2000	NME	0.164	0.556	0.568	0.536	0.583	-9.153	-4.674	-10.121	-5.419	-3.105	0.235	0.892	1.198	-0.458	0.189
24	0.3	300	NME	0.485	-0.151	-0.417	-0.222	0.048	-6.324	-4.184	-10.175	-2.948	-4.514	1.978	-0.963	-1.564	-0.815	-0.798
24	0.3	500	NME	0.359	-0.256	-0.246	-0.528	-0.455	-4.076	-2.770	-6.850	-2.268	-3.354	1.501	-0.657	-0.871	-0.929	-0.947
24	0.3	2000	NME	0.061	-0.723	-0.778	-0.842	-0.768	-1.213	-1.382	-2.646	-1.318	-1.551	0.391	-0.820	-0.942	-0.933	-0.878
48	0.1	100	NME	1.479	4.214	3.596	4.099	4.263	-23.010	-21.144	-30.086	-12.096	-14.898	2.922	1.425	-1.443	7.824	5.053
48	0.1	150	NME	1.507	3.469	3.320	3.199	3.494	-21.734	-18.989	-28.074	-11.759	-13.828	2.098	1.365	-0.997	6.550	4.417
48	0.1	200	NME	0.588	3.061	2.925	2.598	3.009	-21.333	-17.257	-26.378	-11.121	-12.643	0.681	1.595	-0.200	4.785	3.522
48	0.1	250	NME	0.645	2.637	2.438	2.673	2.805	-20.457	-15.935	-25.418	-10.530	-11.622	0.667	1.743	0.464	3.743	2.906
48	0.1	2000	NME	0.100	1.966	2.065	1.660	1.784	-8.652	-2.749	-6.919	-2.743	-2.029	0.144	2.241	2.582	0.916	1.415
48	0.3	2000	NME	-0.267	3.554	3.432	3.661	3.650	-1.684	2.752	1.512	2.644	2.809	0.224	3.624	3.991	3.160	3.339
48	0.5	300	NME	-0.151	10.938	11.888	10.768	10.780	-2.467	7.916	2.273	9.729	8.154	2.582	9.159	7.567	12.796	9.865
48	0.5	500	NME	-0.507	10.479	8.977	10.263	10.389	-2.000	8.621	3.134	9.609	8.762	1.318	9.419	6.431	11.436	9.847
48	0.5	2000	NME	-0.565	10.016	9.748	9.755	9.914	-0.964	9.539	8.246	9.581	9.497	0.422	9.788	9.130	10.074	9.806

Table E2: *RMSE of MMLE, Empirical Bayes, & Hierarchical Bayes by Item Parameter Type for Research Question 1a*

				MMLE					Empirical Bayes					Hierarchical Bayes				
# Items	RSD	# Persons	Q-matrix	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}
24	0.1	150	ME	0.205	0.366	0.253	0.267	0.392	0.396	0.814	0.597	0.220	0.448	0.156	0.276	0.185	0.189	0.278
24	0.1	200	ME	0.177	0.344	0.228	0.240	0.357	0.380	0.732	0.556	0.203	0.417	0.131	0.264	0.172	0.173	0.262
24	0.1	250	ME	0.156	0.322	0.209	0.221	0.334	0.367	0.701	0.546	0.188	0.391	0.118	0.253	0.161	0.163	0.250
24	0.1	300	ME	0.142	0.308	0.195	0.209	0.320	0.355	0.662	0.513	0.180	0.364	0.110	0.248	0.152	0.157	0.246
24	0.1	500	ME	0.110	0.282	0.172	0.180	0.290	0.311	0.518	0.439	0.154	0.291	0.094	0.238	0.140	0.144	0.236
24	0.1	2000	ME	0.054	0.245	0.134	0.140	0.247	0.181	0.267	0.212	0.125	0.215	0.071	0.229	0.124	0.128	0.228
24	0.3	300	ME	0.146	0.616	0.345	0.295	0.601	0.178	0.534	0.303	0.255	0.518	0.155	0.547	0.304	0.255	0.525
24	0.3	500	ME	0.112	0.596	0.323	0.270	0.571	0.133	0.543	0.294	0.244	0.519	0.118	0.553	0.299	0.245	0.524
24	0.3	2000	ME	0.055	0.565	0.294	0.237	0.539	0.060	0.551	0.286	0.228	0.522	0.056	0.554	0.288	0.231	0.527
24	0.5	2000	ME	0.057	0.938	0.375	0.416	0.883	0.059	0.924	0.367	0.411	0.870	0.057	0.931	0.371	0.412	0.875
48	0.1	150	ME	0.197	0.396	0.260	0.281	0.428	0.375	0.677	0.475	0.193	0.379	0.130	0.260	0.165	0.195	0.295
48	0.1	200	ME	0.169	0.367	0.235	0.256	0.393	0.360	0.624	0.448	0.179	0.348	0.119	0.253	0.159	0.187	0.285
48	0.1	250	ME	0.150	0.350	0.218	0.235	0.370	0.350	0.578	0.420	0.169	0.322	0.111	0.252	0.151	0.175	0.276
48	0.1	300	ME	0.138	0.337	0.206	0.223	0.356	0.338	0.536	0.396	0.161	0.302	0.106	0.247	0.145	0.170	0.271
48	0.1	500	ME	0.105	0.307	0.178	0.196	0.326	0.294	0.420	0.326	0.143	0.249	0.090	0.240	0.135	0.159	0.268
48	0.1	2000	ME	0.053	0.270	0.140	0.155	0.283	0.158	0.231	0.142	0.120	0.213	0.053	0.246	0.124	0.144	0.266
48	0.3	500	ME	0.106	0.609	0.326	0.341	0.623	0.127	0.571	0.309	0.318	0.573	0.107	0.571	0.303	0.313	0.576
48	0.3	2000	ME	0.052	0.585	0.301	0.311	0.594	0.062	0.578	0.298	0.303	0.576	0.053	0.576	0.295	0.304	0.582
24	0.1	150	NME	0.204	0.362	0.244	0.259	0.375	0.297	0.582	0.441	0.239	0.402	0.204	0.368	0.246	0.195	0.324
24	0.1	200	NME	0.174	0.333	0.222	0.234	0.350	0.273	0.538	0.417	0.225	0.373	0.178	0.339	0.223	0.182	0.308
24	0.1	250	NME	0.157	0.311	0.203	0.215	0.326	0.264	0.512	0.397	0.208	0.349	0.163	0.316	0.205	0.172	0.291
24	0.1	300	NME	0.144	0.300	0.192	0.202	0.308	0.253	0.479	0.378	0.198	0.331	0.150	0.305	0.195	0.165	0.278
24	0.1	500	NME	0.109	0.272	0.165	0.174	0.281	0.219	0.395	0.324	0.170	0.282	0.113	0.274	0.168	0.150	0.259
24	0.1	2000	NME	0.054	0.230	0.124	0.134	0.237	0.117	0.244	0.183	0.124	0.222	0.055	0.231	0.125	0.128	0.232
24	0.3	300	NME	0.144	0.594	0.362	0.329	0.599	0.162	0.528	0.318	0.284	0.523	0.160	0.562	0.345	0.287	0.543
24	0.3	500	NME	0.113	0.577	0.345	0.311	0.581	0.124	0.535	0.306	0.282	0.532	0.115	0.557	0.333	0.285	0.547
24	0.3	2000	NME	0.056	0.552	0.318	0.284	0.555	0.058	0.540	0.304	0.276	0.541	0.056	0.547	0.315	0.278	0.547
48	0.1	100	NME	0.244	0.439	0.306	0.308	0.442	0.306	0.576	0.426	0.216	0.395	0.221	0.441	0.306	0.210	0.369
48	0.1	150	NME	0.194	0.380	0.260	0.262	0.385	0.280	0.520	0.392	0.193	0.362	0.189	0.385	0.261	0.188	0.335
48	0.1	200	NME	0.168	0.351	0.234	0.233	0.353	0.266	0.479	0.367	0.177	0.332	0.171	0.355	0.235	0.171	0.313
48	0.1	250	NME	0.149	0.332	0.218	0.216	0.333	0.253	0.450	0.353	0.168	0.313	0.154	0.343	0.221	0.162	0.299
48	0.1	2000	NME	0.052	0.248	0.138	0.134	0.246	0.110	0.234	0.157	0.116	0.214	0.053	0.251	0.141	0.126	0.240
48	0.3	2000	NME	0.053	0.487	0.236	0.268	0.503	0.057	0.476	0.228	0.258	0.488	0.053	0.483	0.235	0.260	0.493
48	0.5	300	NME	0.145	0.899	0.476	0.365	0.860	0.148	0.830	0.436	0.336	0.787	0.147	0.839	0.437	0.339	0.799
48	0.5	500	NME	0.112	0.877	0.455	0.343	0.836	0.117	0.835	0.430	0.325	0.792	0.114	0.842	0.431	0.328	0.801
48	0.5	2000	NME	0.055	0.848	0.430	0.314	0.806	0.057	0.838	0.423	0.310	0.795	0.056	0.840	0.424	0.311	0.798

Table E3: *RPB and SDB of Hierarchical Bayes with Item Covariates by Item Parameter Type for Research Question 1b*

				RPB					SDB				
Items	RSD	Persons	Q-matrix	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}
24	0.1	100	ME	6.215	2.329	2.114	0.928	1.568	6.014	3.382	2.700	2.580	4.358
24	0.3	100	ME	7.849	4.169	-4.740	-25.937	5.037	9.564	4.973	6.182	11.148	7.546
24	0.3	150	ME	5.671	4.095	-2.866	-18.354	4.705	8.383	2.835	4.492	6.483	3.133
24	0.3	200	ME	4.523	4.322	-0.958	-15.113	4.266	7.893	5.361	5.857	8.412	5.919
24	0.3	250	ME	3.517	4.289	-0.162	-9.961	4.165	9.860	4.576	5.672	6.013	4.926
24	0.5	100	ME	15.109	-0.708	14.859	1.726	-1.161	3.990	-0.792	3.020	4.645	3.750
24	0.5	150	ME	9.782	-1.092	5.097	0.370	-1.352	2.827	2.079	4.348	5.518	4.841
24	0.5	200	ME	7.130	-1.489	-0.571	0.854	-1.181	5.650	1.653	2.484	3.944	3.331
24	0.5	250	ME	4.879	-1.322	4.465	-0.529	-1.685	-0.752	1.255	4.188	4.021	2.744
24	0.5	300	ME	4.560	-1.382	-0.934	-0.004	-1.382	2.617	1.894	2.687	3.302	2.698
24	0.5	500	ME	2.809	-1.592	-2.040	-0.254	-1.135	0.957	0.427	0.404	1.262	0.727
48	0.1	100	ME	4.740	5.190	5.748	5.370	4.611	12.617	13.992	12.723	11.552	14.500
48	0.3	100	ME	8.998	-3.100	-0.013	-1.553	-2.898	21.294	15.171	14.523	15.754	14.812
48	0.3	150	ME	6.196	-2.935	-0.540	-3.150	-3.460	12.179	9.568	9.749	11.589	10.847
48	0.3	200	ME	5.140	-3.097	-1.308	-2.808	-3.246	16.269	8.957	8.720	11.005	9.981
48	0.3	250	ME	3.813	-2.982	-1.472	-3.048	-3.265	12.829	8.359	7.904	7.948	7.886
48	0.3	300	ME	2.835	-3.077	-1.954	-3.117	-3.266	3.860	3.631	5.926	6.549	7.339
48	0.5	100	ME	15.580	-6.469	14.346	-12.631	-10.247	12.026	8.770	8.327	10.386	10.489
48	0.5	150	ME	9.665	-6.101	12.919	-11.718	-9.556	4.754	2.082	2.825	4.074	3.985
48	0.5	200	ME	6.941	-6.819	3.285	-9.553	-8.351	6.014	4.380	4.988	4.940	5.358
48	0.5	250	ME	5.626	-6.857	2.318	-8.772	-8.015	4.393	4.047	4.273	4.368	4.103
48	0.5	300	ME	4.800	-6.758	-0.838	-9.590	-8.023	4.246	2.589	2.704	2.991	3.400
48	0.5	500	ME	2.519	-6.867	-3.824	-7.480	-7.154	2.694	1.217	1.222	2.247	2.205
48	0.5	2000	ME	0.617	-6.586	-5.520	-6.662	-6.754	0.059	0.612	0.588	0.155	0.504
24	0.1	100	NME	4.722	2.464	0.334	1.944	3.022	11.665	3.832	4.838	5.875	7.583
24	0.3	100	NME	7.735	-0.590	-2.141	1.114	0.237	10.397	4.506	8.035	11.191	10.052
24	0.3	150	NME	5.051	-0.801	-1.955	0.219	-0.166	7.040	5.577	6.891	7.903	7.713
24	0.3	200	NME	3.744	-0.845	-1.741	-0.050	-0.455	6.958	4.284	4.897	4.893	5.653
24	0.3	250	NME	3.097	-0.805	-1.436	-0.611	-0.888	4.345	2.958	2.847	3.234	3.646
24	0.5	100	NME	11.491	2.147	3.035	6.997	3.299	8.659	6.074	5.536	7.402	5.898
24	0.5	150	NME	6.458	2.719	4.175	4.482	2.379	5.220	3.760	3.168	3.989	3.507
24	0.5	200	NME	4.883	2.161	2.810	3.043	2.485	3.237	2.432	3.753	2.982	3.004
24	0.5	250	NME	3.572	2.521	3.318	1.653	2.229	0.053	2.103	3.487	3.166	3.136
24	0.5	300	NME	3.390	2.169	2.467	3.358	2.444	2.467	0.784	1.811	1.297	1.437
24	0.5	500	NME	1.992	2.317	2.245	3.247	2.532	2.329	0.715	1.790	2.820	0.800
24	0.5	2000	NME	0.424	2.514	2.533	1.924	2.248	-0.210	1.159	0.649	-0.059	0.884
48	0.1	300	NME	0.587	2.090	1.260	2.872	2.518	2.146	4.397	4.413	15.612	10.207
48	0.1	500	NME	0.516	2.436	2.536	1.040	1.506	8.719	3.966	4.197	16.711	9.723
48	0.3	100	NME	6.876	2.965	4.968	7.244	4.303	14.646	8.872	10.253	17.197	11.793
48	0.3	150	NME	4.898	4.032	6.972	3.828	3.202	11.846	6.733	7.699	13.308	10.391
48	0.3	200	NME	3.442	4.109	6.451	2.945	2.788	9.202	5.510	6.424	10.653	7.895
48	0.3	250	NME	2.723	3.826	5.861	3.021	2.983	0.994	-0.118	0.872	5.258	4.495
48	0.3	300	NME	2.344	3.997	5.696	2.847	3.063	3.476	0.393	-1.607	-6.605	-5.598
48	0.3	500	NME	1.436	4.198	5.969	1.922	2.771	-6.587	0.875	1.950	3.933	2.593
48	0.5	100	NME	11.102	8.267	-0.105	18.590	10.525	9.130	7.003	6.587	9.786	9.001
48	0.5	150	NME	6.448	8.657	3.103	15.567	10.093	2.738	3.591	4.591	6.422	6.570
48	0.5	200	NME	4.516	8.937	4.173	13.661	10.063	4.959	3.913	4.502	4.919	5.375
48	0.5	250	NME	3.350	9.283	4.813	12.617	9.831	3.725	3.362	3.073	3.446	3.731

Table E4: *RPB of Hierarchical Bayes with/without Item Covariates by Item Parameter Type for Research Question 2*

				Hierarchical Bayes w/Item Covariates					Hierarchical Bayes w/out Item Covariates				
Items	RSD	Persons	Q-matrix	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}
24	0.1	100	ME	6.215	2.329	2.114	0.928	1.568	1.187	2.316	9.627	4.874	0.566
24	0.1	150	ME	4.154	0.911	0.625	1.993	1.713	1.631	1.836	8.041	3.863	0.941
24	0.1	200	ME	3.309	1.303	1.411	0.574	0.806	1.636	2.533	8.046	1.238	-0.005
24	0.1	250	ME	2.43	0.688	0.618	0.926	0.805	1.195	1.939	6.359	1.322	0.125
24	0.1	300	ME	2.031	0.887	1.139	0.255	0.487	1.07	1.996	5.878	0.553	-0.053
24	0.1	500	ME	1.302	0.637	0.874	0.329	0.445	0.845	1.696	4.344	0.039	-0.065
24	0.1	2000	ME	0.309	0.372	0.492	0.133	0.218	0.264	0.786	1.546	0.032	0.112
24	0.3	100	ME	7.849	4.169	-4.74	-25.937	5.037	5.203	3.901	1.242	-45.848	3.454
24	0.3	150	ME	5.671	4.095	-2.866	-18.354	4.705	4.022	4.238	1.878	-30.405	3.731
24	0.3	200	ME	4.523	4.322	-0.958	-15.113	4.266	3.144	4.341	2.603	-23.153	3.618
24	0.3	250	ME	3.517	4.289	-0.162	-9.961	4.165	2.534	4.244	2.696	-15.714	3.757
24	0.3	300	ME	2.904	3.81	-0.799	-6.807	4.59	1.994	3.884	1.797	-11.078	4.171
24	0.3	500	ME	1.712	4.309	1.775	-1.814	4.441	1.221	4.278	3.137	-4.42	4.302
24	0.3	2000	ME	0.374	4.426	3.907	3.192	4.446	0.26	4.417	4.272	2.333	4.336
24	0.5	100	ME	15.109	-0.708	14.859	1.726	-1.161	12.592	-0.642	16.341	0.334	-3.267
24	0.5	150	ME	9.782	-1.092	5.097	0.37	-1.352	8.156	-1.045	7.533	-0.781	-2.859
24	0.5	200	ME	7.13	-1.489	-0.571	0.854	-1.181	5.659	-1.372	1.693	-0.365	-2.515
24	0.5	250	ME	4.879	-1.322	4.465	-0.529	-1.685	4.182	-1.199	6.084	-1.476	-2.639
24	0.5	300	ME	4.56	-1.382	-0.934	-0.004	-1.382	3.681	-1.293	1.183	-0.887	-2.23
24	0.5	500	ME	2.809	-1.592	-2.04	-0.254	-1.135	2.373	-1.488	-0.519	-0.945	-1.725
24	0.5	2000	ME	0.342	-1.532	-0.183	-1.383	-1.503	0.624	-1.437	-0.497	-1.524	-1.627
48	0.1	100	ME	4.74	5.19	5.748	5.37	4.611	1.685	7.331	17.162	5.868	2.545
48	0.1	150	ME	3.295	4.38	4.526	5.878	4.789	1.826	7.497	16.14	3.074	2.317
48	0.1	200	ME	2.792	4.179	4.43	5.47	4.653	1.793	7.484	15.163	1.822	2.2
48	0.1	250	ME	2.063	4.442	5.062	4.385	4.132	1.308	7.701	14.903	0.385	1.775
48	0.1	300	ME	1.998	4.236	4.421	4.85	4.29	1.36	7.465	13.673	0.577	1.943
48	0.1	500	ME	1.271	3.994	4.222	4.661	4.282	0.918	6.618	11.078	1.038	2.43
48	0.1	2000	ME	0.428	4.115	4.158	4.179	4.152	0.288	5.077	6.464	2.894	3.554
48	0.3	100	ME	8.998	-3.1	-0.013	-1.553	-2.898	6.418	-2.791	5.417	-4.725	-5.004
48	0.3	150	ME	6.196	-2.935	-0.54	-3.15	-3.46	4.664	-2.483	3.479	-5.383	-4.773
48	0.3	200	ME	5.14	-3.097	-1.308	-2.808	-3.246	3.756	-2.782	1.551	-4.309	-4.109
48	0.3	250	ME	3.813	-2.982	-1.472	-3.048	-3.265	2.71	-2.755	0.698	-4.139	-3.903
48	0.3	300	ME	2.835	-3.077	-1.954	-3.117	-3.266	1.972	-2.891	-0.305	-3.838	-3.653
48	0.3	500	ME	2.177	-3.043	-2.447	-2.542	-2.809	1.605	-2.938	-1.445	-2.978	-3.051
48	0.3	2000	ME	0.297	-2.744	-2.573	-2.735	-2.75	0.042	-2.735	-2.469	-2.782	-2.801
48	0.5	100	ME	15.58	-6.469	14.346	-12.631	-10.247	12.414	-6.091	30.358	-14.116	-10.83
48	0.5	150	ME	9.665	-6.101	12.919	-11.718	-9.556	7.862	-5.8	23.525	-12.917	-10.597
48	0.5	200	ME	6.941	-6.819	3.285	-9.553	-8.351	5.234	-6.554	11.942	-10.61	-9.343
48	0.5	250	ME	5.626	-6.857	2.318	-8.772	-8.015	4.57	-6.666	8.984	-9.736	-8.69
48	0.5	300	ME	4.8	-6.758	-0.838	-9.59	-8.023	3.849	-6.645	4.245	-10.117	-8.546
48	0.5	500	ME	2.519	-6.867	-3.824	-7.48	-7.154	2.026	-6.772	-0.524	-7.869	-7.501
48	0.5	2000	ME	0.617	-6.586	-5.52	-6.662	-6.754	0.475	-6.517	-4.088	-7.2	-6.896

Table E5: *RMSE of Hierarchical Bayes with/without Item Covariates by Item Parameter Type for Research Question 2*

				Hierarchical Bayes w/Item Covariates					Hierarchical Bayes w/out Item Covariates				
Items	RSD	Persons	Q-matrix	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}
24	0.1	100	ME	0.181	0.303	0.212	0.217	0.301	0.151	0.33	0.292	0.243	0.262
24	0.1	150	ME	0.156	0.276	0.185	0.189	0.278	0.137	0.312	0.263	0.2	0.236
24	0.1	200	ME	0.131	0.264	0.172	0.173	0.262	0.126	0.303	0.238	0.176	0.224
24	0.1	250	ME	0.118	0.253	0.161	0.163	0.25	0.118	0.291	0.219	0.16	0.218
24	0.1	300	ME	0.11	0.248	0.152	0.157	0.246	0.113	0.284	0.205	0.154	0.219
24	0.1	500	ME	0.094	0.238	0.14	0.144	0.236	0.094	0.266	0.175	0.139	0.222
24	0.1	2000	ME	0.071	0.229	0.124	0.128	0.228	0.052	0.24	0.134	0.126	0.229
24	0.3	100	ME	0.26	0.55	0.334	0.294	0.534	0.241	0.511	0.338	0.28	0.484
24	0.3	150	ME	0.218	0.551	0.322	0.275	0.531	0.203	0.53	0.326	0.263	0.502
24	0.3	200	ME	0.191	0.548	0.313	0.261	0.52	0.182	0.537	0.32	0.257	0.506
24	0.3	250	ME	0.167	0.547	0.306	0.257	0.521	0.161	0.539	0.312	0.255	0.513
24	0.3	300	ME	0.155	0.547	0.304	0.255	0.525	0.149	0.541	0.309	0.253	0.519
24	0.3	500	ME	0.118	0.553	0.299	0.245	0.524	0.115	0.549	0.302	0.245	0.522
24	0.3	2000	ME	0.056	0.554	0.288	0.231	0.527	0.056	0.555	0.288	0.231	0.525
24	0.5	100	ME	0.271	0.948	0.435	0.449	0.843	0.257	0.892	0.422	0.427	0.799
24	0.5	150	ME	0.216	0.938	0.415	0.439	0.852	0.209	0.904	0.409	0.425	0.822
24	0.5	200	ME	0.18	0.934	0.405	0.433	0.859	0.184	0.914	0.402	0.423	0.837
24	0.5	250	ME	0.169	0.936	0.397	0.43	0.863	0.162	0.917	0.394	0.421	0.844
24	0.5	300	ME	0.148	0.935	0.395	0.428	0.865	0.147	0.92	0.392	0.421	0.85
24	0.5	500	ME	0.114	0.933	0.385	0.426	0.876	0.114	0.924	0.384	0.422	0.865
24	0.5	2000	ME	0.057	0.931	0.371	0.412	0.875	0.057	0.929	0.37	0.412	0.873
48	0.1	100	ME	0.155	0.28	0.185	0.211	0.304	0.155	0.329	0.279	0.228	0.249
48	0.1	150	ME	0.13	0.26	0.165	0.195	0.295	0.138	0.313	0.25	0.185	0.239
48	0.1	200	ME	0.119	0.253	0.159	0.187	0.285	0.128	0.305	0.231	0.17	0.24
48	0.1	250	ME	0.111	0.252	0.151	0.175	0.276	0.12	0.301	0.217	0.159	0.241
48	0.1	300	ME	0.106	0.247	0.145	0.17	0.271	0.113	0.295	0.205	0.152	0.242
48	0.1	500	ME	0.09	0.24	0.135	0.159	0.268	0.092	0.282	0.179	0.145	0.252
48	0.1	2000	ME	0.053	0.246	0.124	0.144	0.266	0.05	0.265	0.142	0.139	0.263
48	0.3	100	ME	0.215	0.559	0.335	0.334	0.556	0.237	0.521	0.33	0.31	0.508
48	0.3	150	ME	0.19	0.562	0.323	0.326	0.558	0.196	0.536	0.319	0.311	0.53
48	0.3	200	ME	0.164	0.561	0.315	0.321	0.562	0.172	0.543	0.312	0.31	0.544
48	0.3	250	ME	0.149	0.564	0.313	0.319	0.566	0.153	0.549	0.31	0.31	0.553
48	0.3	300	ME	0.142	0.568	0.31	0.317	0.567	0.14	0.554	0.308	0.311	0.557
48	0.3	500	ME	0.107	0.571	0.303	0.313	0.576	0.107	0.565	0.302	0.309	0.571
48	0.3	2000	ME	0.053	0.576	0.295	0.304	0.582	0.053	0.574	0.294	0.303	0.58
48	0.5	100	ME	0.255	0.884	0.41	0.452	0.803	0.261	0.874	0.417	0.437	1.364
48	0.5	150	ME	0.211	0.917	0.411	0.46	0.84	0.212	0.909	0.411	0.445	0.81
48	0.5	200	ME	0.177	0.925	0.403	0.465	0.86	0.188	0.922	0.406	0.454	0.838
48	0.5	250	ME	0.159	0.933	0.403	0.464	0.866	0.16	0.928	0.403	0.456	0.849
48	0.5	300	ME	0.143	0.938	0.401	0.467	0.879	0.143	0.934	0.401	0.461	0.865
48	0.5	500	ME	0.11	0.952	0.399	0.47	0.895	0.11	0.95	0.399	0.465	0.885
48	0.5	2000	ME	0.054	0.966	0.393	0.47	0.911	0.053	0.967	0.393	0.469	0.909

Table E6: *SDB of Empirical Bayes and Hierarchical Bayes by Item Parameter Type for Research Question 3*

				Empirical Bayes					Hierarchical Bayes				
Items	RSD	Persons	Q-matrix	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}
24	0.1	150	ME	-1.322	-6.472	-10.202	4.114	5.339	-0.837	2.254	4.333	3.941	2.851
24	0.1	200	ME	-2.223	-10.249	-12.988	3.996	6.155	7.408	6.017	6.043	4.501	6.452
24	0.1	250	ME	-2.017	-10.192	-17.64	4.182	3.795	10.245	8.672	8.42	8.513	8.849
24	0.1	300	ME	-3.682	-14.543	-13.297	3.543	3.83	10.933	7.604	11.448	8.098	9.115
24	0.1	500	ME	-8.441	-6.633	-14.387	4.407	4.638	9.598	8.914	10.601	8.126	9.048
24	0.1	2000	ME	-6.107	2.973	-4.779	3.621	6.012	-23.86	0.179	2.999	2.322	2.288
24	0.3	300	ME	6.029	5.551	2.453	2.766	5.052	7.037	3.715	3.415	5.298	4.358
24	0.3	500	ME	4.136	3.876	2.058	3.254	4.257	3.963	2.388	2.275	4.692	3.834
24	0.3	2000	ME	3.269	1.699	0.813	1.857	1.448	1.832	1.026	0.378	0.776	0.659
24	0.5	2000	ME	0.207	0.883	1.173	0.389	0.901	-0.846	0.04	0.644	0.026	0.2
48	0.1	150	ME	23.822	29.738	14.627	18.61	27.652	14.272	15.775	14.574	12.107	12.832
48	0.1	200	ME	22.579	23.756	11.953	15.127	26.812	14.04	13.014	9.324	8.19	12.49
48	0.1	250	ME	20.382	17.192	7.859	14.423	23.463	12.828	11.276	10.751	10.736	13.328
48	0.1	300	ME	16.372	14.616	7.097	13.78	21.265	14.756	13.645	13.261	13.939	14.982
48	0.1	500	ME	11.08	13.998	6.541	13.018	15.967	22.427	12.677	12.844	13.574	11.926
48	0.1	2000	ME	7.971	10.044	6.165	7.382	9.95	10.557	6.278	6.215	6.122	5.902
48	0.3	500	ME	8.336	6.878	4.586	3.216	6.432	7.801	3.964	4.909	4.643	5.039
48	0.3	2000	ME	3.641	2.994	1.613	1.31	2.143	1.895	1.75	1.244	1.018	1.021
24	0.1	150	NME	7.704	2.632	-0.893	7.226	13.233	10.688	4.07	6.227	11.016	9.128
24	0.1	200	NME	8.313	1.405	-4.833	6.25	10.305	11.249	2.625	3.025	11.604	7.215
24	0.1	250	NME	4.735	-1.438	-5.359	7.605	11.922	7.838	3.19	4.588	11.09	6.667
24	0.1	300	NME	2.362	-0.97	-6.629	6.802	13.898	6.442	1.838	3.183	11.704	8.645
24	0.1	500	NME	1.229	0.863	-7.203	7.03	10.455	6.202	2.464	1.38	9.867	5.217
24	0.1	2000	NME	1.578	3.648	-2.291	4.5	5.111	1.524	-0.368	0.091	3.039	1.251
24	0.3	300	NME	10.574	5.225	5.506	4.786	8.383	-4.569	-2.717	-0.997	2.905	2.37
24	0.3	500	NME	5.101	3.066	3.223	2.693	3.841	1.653	1.044	0.707	2.254	1.498
24	0.3	2000	NME	1.222	1.042	1.923	0.95	0.342	-0.039	0.389	0.951	0.75	-0.385
48	0.1	100	NME	24.216	27.252	15.9	18.362	28.358	17.239	9.698	10.967	11.559	13.657
48	0.1	150	NME	24.851	25.268	15.463	18.296	25.791	19.454	9.507	10.839	16.582	14.044
48	0.1	200	NME	22.758	21.573	12.703	18.47	24.302	7.91	6.68	7.566	18.069	12.052
48	0.1	250	NME	20.988	17.359	7.964	15.985	19.87	7.571	0.218	4.718	16.292	11.286
48	0.1	2000	NME	6.871	7.363	4.908	6.599	7.563	1.171	1.354	1.595	5.724	3.926
48	0.3	2000	NME	2.184	1.61	1.55	2.009	2.059	0.64	0.78	0.829	1.676	1.362
48	0.5	300	NME	5.495	4.389	2.137	1.992	5.26	3.222	1.73	1.921	1.826	2.101
48	0.5	500	NME	2.638	1.715	2.627	1.907	3.555	1.068	0.104	2.227	1.605	1.492
48	0.5	2000	NME	1.441	1.864	0.127	-0.315	0.47	0.98	1.337	-0.099	-0.614	0.017

Appendix F

Method Selection Guideline Supplement

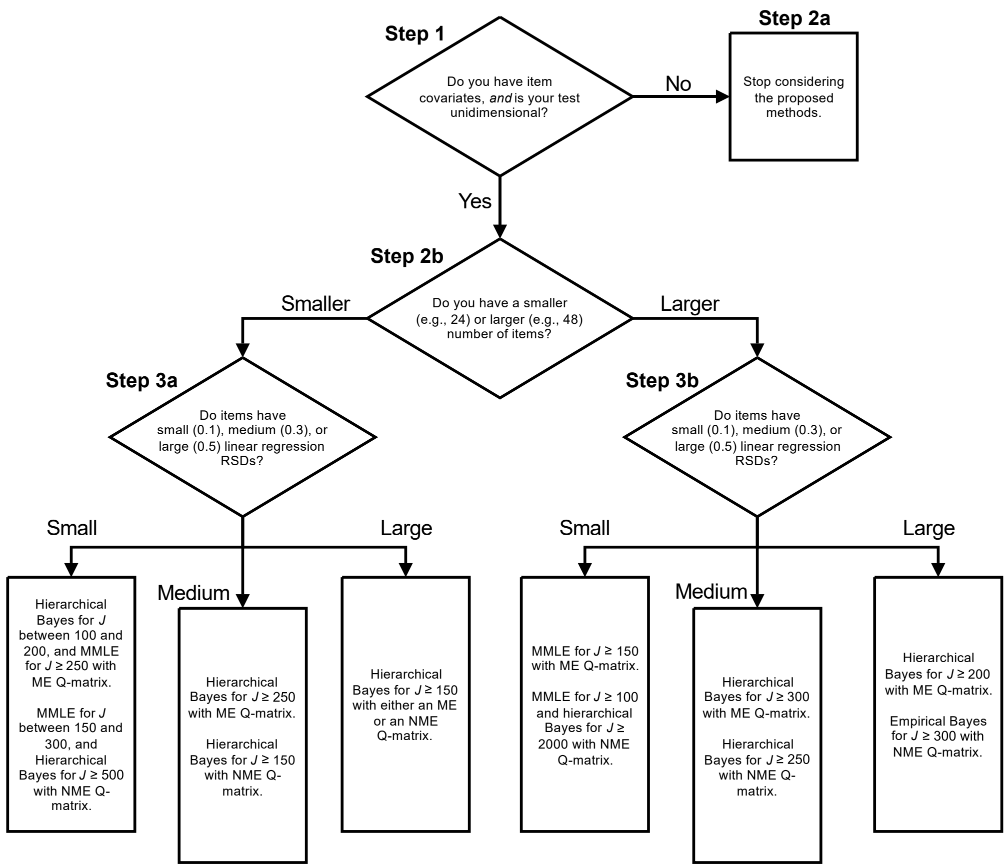

Figure F1: Method selection guideline

References

Baker, F. B. (1993). Sensitivity of the linear logistic test model to misspecification of the weight matrix. *Applied Psychological Measurement, 17*(3), 201–210. https://doi.org/10.1177/014662169301700301

Bechger, T. M., Verstralen, H. H. F. M. & Verhelst, N. D. (2002). Equivalent linear logistic test models. *Psychometrika, 67*, 123–136. https://doi.org/10.1007/BF02294712

Beretvas, S. N., & Williams, N. J. (2004). The use of hierarchical generalized linear model for item dimensionality assessment. *Journal of Educational Measurement, 41*(4), 379-395. https://doi.org/10.1111/j.1745-3984.2004.tb01172.x

Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement, 39*(4), 331-348. https://doi.org/10.1111/j.1745-3984.2002.tb01146.x

Chalmers, R. P. (2015). Extended mixed-effects item response models with the MH-RM algorithm. *Journal of Educational Measurement, 52*(2), 200-222. https://doi.org/10.1111/jedm.12072

Cho, S.-J., Li, F., & Bandalos, D. L. (2009). Accuracy of the parallel analysis procedure using polychoric correlations. *Educational and Psychological Measurement, 69*, 748-759.

Choi, I. H., & Wilson, M. (2015). Multidimensional classification of examinees using the mixture random weights linear logistic test model. *Educational and psychological measurement, 75*(1), 78-101. https://doi.org/10.1177%2F0013164414522124

Davis, H., Schlundt, D., Camarata, S., Bess, F., & Hornsby, B. (2020, in review). Understanding listening-related fatigue: Perspectives of adults with hearing loss. *International Journal of Audiology.*

Embretson, S. E. (2015). The multicomponent latent trait model for diagnosis: Applications to heterogeneous test domains. *Applied Psychological Measurement, 39*(1), 16-30. https://doi.org/10.1177%2F0146621614552014

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*(6), 359-374. https://doi.org/10.1016/0001-6918(73)90003-6

Freund, P. A., Hofer, S., & Holling, H. (2008). Explaining and controlling for the psychometric properties of computer-generated figural matrix items. *Applied Psychological Measurement, 32*(3), 195-210. https://doi.org/10.1177%2F0146621607306972

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–511. http://doi.org/10.1214/ss/1177011136

Gorin, J. S. (2005). Manipulating processing difficulty of reading comprehension questions: The feasibility of verbal item generation. *Journal of Educational Measurement, 42*(4), 351-373. https://doi.org/10.1111/j.1745-3984.2005.00020.x

Hartig, J., Frey, A., Nold, G., & Klieme, E. (2012). An application of explanatory item response modeling for model-based proficiency scaling. *Educational and Psychological Measurement, 72*(4), 665-686. https://doi.org/10.1177%2F0013164411430707

Hoffman, L., Yang, X., Bovaird, J. A., & Embretson, S. E. (2006). Measuring attentional ability in older adults: Development and psychometric evaluation of DriverScan. *Educational and Psychological Measurement, 66*(6), 984-1000. https://doi.org/10.1177%2F0013164406288170

Hohensinn, C., & Kubinger, K. D. (2011). Applying item response theory methods to examine the impact of different response formats. *Educational and Psychological Measurement, 71*(4), 732-746. https://doi.org/10.1177%2F0013164410390032

Hornke, L. F., & Habon, M. W. (1986). Rule-based item bank construction and evaluation within the linear logistic framework. *Applied Psychological Measurement, 10*(4), 369-380. https://doi.org/10.1177%2F014662168601000405

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. http://doi.org/10.1080/10705519909540118

Ip, E. H., Magee, M. F., Youssef, G. A., & Chen, S. H. (2019). Gleaning information for cognitive operations from don't know responses in cognitive and noncognitive assessments. *Multivariate Behavioral Research, 54*(2), 159-172. https://doi.org/10.1080/00273171.2018.1503075

Ip, E. H., Smits, D. J. M., & De Boeck, P. (2009). Locally dependent linear logistic test model with person covariates. *Applied Psychological Measurement, 33*(7), 555-569. https://doi.org/10.1177/0146621608326424

Kim, J. (2018). *Extensions and applications of item explanatory models to polytomous data in item response theory* (Doctoral dissertation, UC Berkeley). Retrieved from https://escholarship.org/uc/item/4pf2p2fj

Kubinger, K. D. (2008). Applications of the linear logistic test model in psychometric research. *Educational and Psychological Measurement, 69*(2), 232–244. https://doi.org/10.1177/0013164408322021

Medina-Díaz, M. (1993). Analysis of cognitive structure using the linear logistic test model and quadratic assignment. *Applied Psychological Measurement, 17*(2), 117-130. https://doi.org/10.1177%2F014662169301700202

Mitchell, K. J. (1983). *Cognitive processing determinants of item difficulty on the verbal subtests of the armed services vocational aptitude battery.* Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Poinstingl, H. (2009). The linear logistic test model (LLTM) as the methodological foundation of item generating rules for a new verbal reasoning test. *Psychology Science Quarterly, 51*(2), 123-134.

Rakkapao, S., Prasitpong, S., & Arayathanitkul, K. (2016). Analysis test of understanding of vectors with the three-parameter logistic model of item response theory and item response curves technique. *Physical Review Physics Education Research, 12*(2), 020135. https://doi.org/10.1103/PhysRevPhysEducRes.12.020135

Rost, J., & Carstensen, C. H. (2002). Multidimensional Rasch measurement via item component models and faceted designs. *Applied Psychological Measurement, 26*(1), 42-56. https://doi.org/10.1177%2F0146621602026001003

Sheehan, K. M., & Mislevy, R. J. (1990). Integrating cognitive and psychometric models to measure document literacy. *Journal of Educational Measurement, 27*(3), 255–272. https://doi.org/10.1111/j.1745-3984.1990.tb00747.x

Shermis, M. D., & Chang, S.-H. (1997). The use of item response theory (IRT) to investigate the hierarchical nature of a college mathematics curriculum. *Educational and Psychological Measurement, 57*(3), 450-458. https://doi.org/10.1177%2F0013164497057003006

Whitely, S. E., & Schneider, L. M. (1981). Information structure for geometric analogies: A test theory approach. *Applied Psychological Measurement, 5*(3), 383-397. https://doi.org/10.1177%2F014662168100500312

Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Doctoral dissertation, University of California, Los Angeles). Retrieved from https://www.statmodel.com/download/Yudissertation.pdf