Multidimensional ability

April 29, 2008

Prepared by:
Murray Aitkin and Irit Aitkin
School of Behavioural Science
University of Melbourne

Prepared for:

US Department of Education
Office of Educational Research and Improvement
National Center for Education Statistics

This project was an activity of the NAEP Education Statistics Services Institute.

Contents

1 Aim of the project

The aim of the project was to investigate the need for simultaneous estimation of correlated abilities on several latent dimensions in NAEP analysis. This report examined models with two and three latent ability dimensions (referred to as factors).

2 Summary

In simulations from two-factor models with correlated factors,

1. fitting a single factor model did not give acceptable reporting group parameter estimates except for highly correlated factors;
2. fitting uncorrelated factors, with or without the correct loading restrictions, gave very good estimates, closely equivalent in bias and precision to those for the correct model;
3. the failure to account for the inter-factor correlation in the two-factor models did not perceptibly affect the properties of the parameter estimates for these models;
4. what matters is that the two-factor structure was recognised: fitting independent factors was simple and effective compared with the additional estimation of the inter-factor correlation.

With real NAEP math data with three subscales,

> the maximized log-likelihood increased dramatically from one through three factors, but the corresponding parameter estimates changes were relatively small, standard error changes were very small.

We conclude that fitting multiple uncorrelated factors with unrestricted loadings appears to give reporting group estimates which are unbiased and almost as precise as the fully efficient estimates assuming scale purity of the items. A computational price is paid for this relative to the estimation of factors with non-zero loadings only for their own scale items; the benefit is in allowing for non-pure items or items whose scale membership is unclear.

3 Simulation studies

We carried out a series of large-scale simulation studies of a two-factor model for ability on 20 binary items. The studies compared the correct two-dimensional model with several simpler models in which either the inter-factor correlation was ignored, or the multidimensionality was ignored.

Ten items were loaded on each factor in 2PL models, using the item parameters from previous simulation studies of the 2PL and other models. The correlation ρ of the two factors was varied from 0 to 0.9 across the studies. Demographic factors were incorporated into the model by a regression function with group factors sex, ethnic origin (4 levels), poverty (2 levels) and homework (3 levels). This regression function was modeled on the logit scale of item responses (the 2PL model), not on the ability distribution (the MIMIC model), so group differences are the same for each ability dimension. (For the MIMIC model it would be possible for group differences to vary by ability dimension. We comment on other difficulties of this model with the NAEP data below.) Parameter values for the group variables were the same as those used in previous studies (given in the tables below).

Four models were fitted to each data set, for each of six factor correlation values ρ: 0, 0.1, 0.3, 0.5, 0.7, and 0.9. The models were:

1. a single factor 2PL model for all 20 items;

2. a correlated two-factor model with 10 loadings constrained to zero on each factor, corresponding to the generating model (the true model);

3. an uncorrelated two-factor model with loadings constrained to zero corresponding to the generating model;

4. an uncorrelated unconstrained two-factor model.

We generated 550 samples of size 1000 from the true model (2) and fitted all four models in LatentGold 4.5.

We encountered considerable difficulties with LatentGold. For about 10% of the samples no output at all was obtained – there was no output file created. In other samples the output file was truncated, with iteration details, and sometimes parameter estimates, missing, though the fitting was successful. There were a few cases of flat likelihoods or sample estimates more than 4 SEs away from the true values; these samples were discarded. Extraction of results from the output was a very tedious and time-consuming process since the format of this file varied with the amount of output.

The results of the four model analyses were compared across the retained samples, in both maximized likelihoods and the bias and mean square error of the parameter estimates. These are shown in the Tables below. The log-likelihood averages are not strictly comparable as they are based on different numbers of samples, but nevertheless show that model 1 was very much inferior to the others for small ρ, but approached the others as ρ increased. Model 2 was generally best but models 3 and 4 were sometimes close, allowing for the one fewer parameter in model 3 and the additional 19 parameters in model 4.

4 Results

1. The single-factor model estimates had substantial biases for the large parameters for small ρ. Their mean square errors were always the largest, though they decreased with increasing ρ, and for large ρ were equivalent to the other methods (since the two factors were then measuring nearly the same ability).

2. The true model 2 nearly always gave the best estimates and mean square errors.

3. The two-factor models 3 and 4 in which the factor correlation was ignored were almost as good as the true model, and were marginally better in a few cases. Their biases were very small, and their standard errors were equivalent to those of the true model.

4. The package SEs (column "se") of the estimates under model 1 were consistently too small because of the model mis-specification; the true variability across samples (column "sdb") was comparable to those of the other method estimates, which were slightly underestimated by the package SE report.

5. The SEs for the unconstrained model 4 were very close to those for the constrained model 3: ignoring the loading constraint negligibly affected the precisions of the estimates. A few samples gave standard errors of 1000 for one of the method 4 model parameters, because the model was unidentifiable or the fitting procedure had failed; these cases were averaged over and give unusually large average SEs.

5 NAEP analysis

We analysed all 79 items of the full 1986 NAEP math test (Grade 3/Age 9), using one, two and three-factor models, and a three-level model accounting for school clustering. (The one-factor 2PL model was also fitted with the four-level model accounting for both school and PSU clustering. Changes in parameter estimates in Table 3 were very small, as with the 30-item data in earlier reports.)

The test is made up of three scales: Measurement with 26 items, Numbers and Operations – High Level with 23 items, and Numbers and Operations – Knowledge and Skills, with 30 items, but we do not distinguish the scales in any of the analyses: all items are treated in common across the single or multiple factors fitted. We fitted a common reporting group structure across all factors; this was done on the logit scale for the 2PL models. For the MIMIC model only the one-factor model was fitted, for reasons explained in the Appendix.

Since the loadings are arbitrary up to an orthogonal rotation for the multiple-factor model, we do not present estimates of the item parameters, only the reporting group estimates and SEs. The ability factor variances are set to 1 and the inter-factor correlations are set to zero; this greatly speeds up the analysis.

6 Results

For the null models with no reporting group variables, and the full reporting group regression models, the maximized log-likelihoods are given in the table below.

```
Maximized log-likelihoods, NAEP 79-item data

Factors  Model     2PL         MIMIC
----------------------------------------
   1      null  -103,377.93  -103,036.12
          full  -103,013.51  -102,562.07
   2      null  -102,557.28
          full  -102,219.51
   3      null  -102,366.59
          full  -102,014.57
----------------------------------------
```

Parameter estimates and standard errors for the models are given in Tables 3 and 4, separately for the 2PL and MIMIC models, for which the parameter estimates are not comparable. (The MIMIC estimates and SEs are much larger, but are closely proportional to the 2PL estimates in most cases, except for the school variance component, which is 11% of the total variance compared to 5-6% for the 2PL models.)

For the 2PL models, the successive improvements in maximized log-likelihood with increasing numbers of factors were 794.00 (1 to 2) and 204.94 (2 to 3). These changes are far beyond any critical value for $\frac{1}{2}\chi^2_{30}$, clearly indicating a three-dimensional ability factor, as expected from the design of the scales. The very large improvements in fit did not correspond to major changes in parameter estimates: these changes were less than one SE except for those for parents education, where two changes of 1.5 SEs occurred. Standard errors increased only slightly with the additional factors.

A notable feauture of the full set of items is that the MIMIC model gave a much better fit than the 2PL, in contradiction to the results for the 30-item Numbers and Operations – Knowledge and Skills scale, where the 2PL model gave a much better fit. For the null model the improvement over the 2PL was 341.81, and this increased to 446.44 for the full model. However we were unable to decide how to deal with the school random effects in the MIMIC model with more than one factor, as discussed in the Appendix.

7 Conclusion

From the simulations it is clear that true multiple ability dimensions need to be recognised in the analysis. The failure to do so, by fitting a single dimension model, led to substantial biases in the large reporting group parameters except when the inter-factor correlation was very high.

Our experience with the NAEP data bore out the simulation results though the biases in the single-factor reporting group parameter estimates were not serious except for the parents education variable. Standard errors were not seriously underestimated by the single-factor NAEP analysis.

Computational times for the three-factor models were substantial, reflecting the large number of parameters in the general three-factor model. The gain in generality with the full model is offset by the heavier computational burden – with three factors the constrained model would have 2x79 item parameters, whereas the general model has 4x79 item parameters!

8 Appendix

8.1 Why does the unconstrained model do so well?

It may be puzzling that the unconstrained uncorrelated factor model does so well compared to the true model. This is because of the rotational invariance of the general factor model which also applies to the binary response factor model. Write the 2PL two-factor model in the form

$$\operatorname{logit} p_{ij} = \boldsymbol{\beta}' \mathbf{x}_i + \lambda_{0j} + \lambda_{1j} z_{1i} + \lambda_{2j} z_{2i},$$

where z_1 and z_2 are the two correlated factors with correlation ρ, λ_{0j} are the item intercepts, and λ_{1j} and λ_{2j} are the loadings of the items on the two factors. The orthogonal rotation

$$u_1 = (z_1 + z_2)/\sqrt{2},\ u_2 = (z_1 - z_2)/\sqrt{2}$$

gives independent factors u_1 and u_2, with

$$z_1 = (u_1 + u_2)/\sqrt{2},\ z_2 = (u_1 - u_2)/\sqrt{2},$$

and the factor model becomes

$$\begin{aligned} \operatorname{logit} p_{ij} &= \beta' \mathbf{x}_i + \lambda_{0j} + \lambda_{1j}(u_{1i} + u_{2i})/\sqrt{2} + \boldsymbol{\lambda}_2(u_1 - u_2)/\sqrt{2} \\ &= \boldsymbol{\beta}' \mathbf{x}_i + \lambda_{0j} + (\lambda_{1j} + \lambda_{2j})u_{1i}/\sqrt{2} + (\lambda_{1j} - \lambda_{2j})u_{2i}/\sqrt{2} \\ &= \boldsymbol{\beta}' \mathbf{x}_i + \lambda_{0j} + \gamma_{1j} u_{1i} + \gamma_{2j} u_{2i}. \end{aligned}$$

If $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ have complementary blocks of zeros, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ will have no zeros; though they are related to smaller-dimension loadings, they can be estimated unbiasedly, though not fully efficiently, by a general two-factor analysis of all the items. This analysis has the additional benefit of guarding against failure of the assumed scale "purity" of the items, and the assumption that they load on only one factor. Thus items can be included in the test even if they load on more than one dimension, or if their loading pattern is uncertain.

8.2 How does the MIMIC model deal with the reporting group regression and the clustering?

Random effects in the 2PL model are placed on the logit scale, with the ability distribution homogeneous. In the MIMIC model they are "inside" the ability distribution model. With more than one dimension of ability, it becomes unclear how the school clustering should be modeled. Writing the two-level two-factor general model out in full, we have

$$\begin{aligned} \operatorname{logit} p_{ij} &= \lambda_{0j} + \lambda_{1j} z_{1i} + \lambda_{2j} z_{2i} \\ (z_{1i}, z_{2i}) &\sim N_2(\boldsymbol{\mu}_i, \Sigma) \end{aligned}$$

where Σ is the random effect covariance matrix with diagonals 1 and off-diagonal ρ and $\boldsymbol{\mu} = (\mu_1, \mu_2)$ is the mean vector.

How does the reporting group regression appear in the two ability means? As we noted in the text, it is quite possible for the two factors to have different regressions $\mu_{1i} = \boldsymbol{\beta}_1' \mathbf{x}_i$ and $\mu_{2i} = \boldsymbol{\beta}_2' \mathbf{x}_i$; to prevent this the two regressions would have to be constrained to be equal. A common regression is natural in the 2PL model where the logit scale for the item response probability holds both the reporting group regression model and the two ability factors.

A more serious difficulty occurs with inclusion of the school random effect for the clustering of students in schools. Indexing schools by k, the three-level two-factor MIMIC model can be written

$$\begin{aligned} \operatorname{logit} p_{ijk} &= \lambda_{0j} + \lambda_{1j} z_{1ik} + \lambda_{2j} z_{2ik} \\ (z_{1ik}, z_{2ik}) &\sim N_2(\boldsymbol{\mu}_i + \boldsymbol{\eta}_k, \Sigma) \\ \boldsymbol{\eta}_k &\sim N(\mathbf{0}, \Psi), \end{aligned}$$

where the school effect $\boldsymbol{\eta} = (\eta_1, \eta_2)$ is now bivariate, like the ability factors, and is in general different on the two ability factors. To constrain the school effect factors to be the same on each ability factor – that is, to require that the effect of the common school environment of students is identical on both ability dimensions – would require the school effect factors (η_1, η_2) to be correlated 1.0 in the model specification, or some other specific model assumption for them. While this is achievable, it requires model validation to assess whether this is a reasonable assumption. We did not perform this investigation, and consider that the MIMIC model does not lend itself easily to multidimensional abilities in clustered designs.

Again, in the 2PL model these effects are on the logit item response scale, and naturally form a single dimension regardless of the number of ability factors.

9 Tables

Table 1

Mean log-likelihood across samples for four methods, simulation

rho	method	mean log-like	samples
0	1	-11479.05	496
	2	-11074.55	524
	3	-11108.84	495
	4	-11048.21	541
0.1	1	-11455.81	534
	2	-11074.86	515
	3	-11077.87	530
	4	-11056.89	521
0.3	1	-11332.11	474
	2	-11039.15	527
	3	-11080.88	521
	4	-11039.50	515
0.5	1	-11222.51	477
	2	-11013.49	523
	3	-11088.79	497
	4	-10999.77	486
0.7	1	-11064.81	509
	2	-10940.52	519
	3	-11110.07	382
	4	-10956.44	503
0.9	1	-10898.97	473
	2	-10830.28	528
	3	-11077.64	420
	4	-10822.48	512

Table 2

Parameter estimates, biases, SEs and MSEs, by rho, simulation

rho	true	mean	bias	mse	sdb	se	param	method
0	-0.472	-0.4096	0.0622	0.0079	0.0635	0.0380	sex	1
	-0.472	-0.4703	0.0015	0.0028	0.0530	0.0514	sex	2
	-0.472	-0.4629	0.0089	0.0091	0.0951	0.0515	sex	3
	-0.472	-0.4719	-0.0001	0.0027	0.0519	0.0515	sex	4
	-0.800	-0.6983	0.1017	0.0176	0.0850	0.0603	pov	1
	-0.800	-0.8011	-0.0011	0.0068	0.0822	0.0809	pov	2
	-0.800	-0.8005	-0.0005	0.0067	0.0817	0.0810	pov	3
	-0.800	-0.8060	-0.0060	0.0067	0.0816	0.0811	pov	4
	0.100	0.0857	-0.0143	0.0044	0.0646	0.0438	hw1	1
	0.100	0.0987	-0.0013	0.0037	0.0612	0.0593	hw1	2
	0.100	0.0967	-0.0033	0.0038	0.0616	0.0593	hw1	3
	0.100	0.0973	-0.0027	0.0038	0.0616	0.0595	hw1	4
	0.300	0.2641	-0.0359	0.0073	0.0777	0.0551	hw2	1
	0.300	0.3009	0.0009	0.0057	0.0758	0.0746	hw2	2
	0.300	0.3001	0.0001	0.0059	0.0766	0.0747	hw2	3
	0.300	0.3022	0.0022	0.0059	0.0765	0.0748	hw2	4
	-2.359	-2.0717	0.2872	0.0921	0.0982	0.0743	ethnic1	1
	-2.359	-2.3640	-0.0052	0.0095	0.0974	0.0968	ethnic1	2
	-2.359	-2.3618	-0.0029	0.0091	0.0954	0.0969	ethnic1	3
	-2.359	-2.3707	-0.0118	0.0098	0.0981	0.0971	ethnic1	4
	-1.887	-1.6500	0.2371	0.0655	0.0963	0.0670	ethnic2	1
	-1.887	-1.8897	-0.0026	0.0094	0.0971	0.0888	ethnic2	2
	-1.887	-1.8887	-0.0016	0.0092	0.0960	0.0888	ethnic2	3
	-1.887	-1.8965	-0.0094	0.0099	0.0988	0.0892	ethnic2	4
	0.944	0.8193	-0.1243	0.0283	0.1135	0.0825	ethnic3	1
	0.944	0.9412	-0.0023	0.0137	0.1170	0.1113	ethnic3	2
	0.944	0.9394	-0.0042	0.0134	0.1158	0.1115	ethnic3	3
	0.944	0.9439	0.0004	0.0135	0.1164	0.1114	ethnic3	4

rho	true	mean	bias	mse	sdb	se	param	method
0.1	-0.472	-0.4223	0.0495	0.0056	0.0563	0.0438	sex	1
	-0.472	-0.4722	-0.0004	0.0030	0.0546	0.0523	sex	2
	-0.472	-0.4717	0.0000	0.0030	0.0548	0.0514	sex	3
	-0.472	-0.4731	-0.0014	0.0032	0.0562	0.0527	sex	4
	-0.800	-0.7169	0.0831	0.0141	0.0850	0.0691	pov	1
	-0.800	-0.8000	-0.0000	0.0069	0.0831	0.0822	pov	2
	-0.800	-0.8013	-0.0013	0.0070	0.0835	0.0809	pov	3
	-0.800	-0.8042	-0.0042	0.0069	0.0832	0.0825	pov	4
	0.100	0.0876	-0.0124	0.0044	0.0650	0.0504	hw1	1
	0.100	0.0991	-0.0009	0.0040	0.0635	0.0602	hw1	2
	0.100	0.0993	-0.0007	0.0039	0.0621	0.0592	hw1	3
	0.100	0.0995	-0.0005	0.0039	0.0627	1.9796	hw1	4
	0.300	0.2738	-0.0262	0.0072	0.0806	0.0634	hw2	1
	0.300	0.3053	0.0053	0.0057	0.0756	0.0759	hw2	2
	0.300	0.3066	0.0066	0.0058	0.0760	0.0745	hw2	3
	0.300	0.3074	0.0075	0.0061	0.0780	1.9955	hw2	4
	-2.359	-2.1011	0.2578	0.0778	0.1065	0.0835	ethnic1	1
	-2.359	-2.3485	0.0104	0.0098	0.0985	0.0980	ethnic1	2
	-2.359	-2.3479	0.0110	0.0098	0.0986	0.0964	ethnic1	3
	-2.359	-2.3564	0.0025	0.0094	0.0972	0.0988	ethnic1	4
	-1.887	-1.6816	0.2055	0.0529	0.1034	0.0766	ethnic2	1
	-1.887	-1.8821	0.0050	0.0083	0.0909	0.0899	ethnic2	2
	-1.887	-1.8801	0.0070	0.0083	0.0909	0.0885	ethnic2	3
	-1.887	-1.8901	-0.0029	0.0085	0.0924	2.0098	ethnic2	4
	0.944	0.8485	-0.0950	0.0256	0.1288	0.0947	ethnic3	1
	0.944	0.9527	0.0091	0.0146	0.1203	0.1134	ethnic3	2
	0.944	0.9498	0.0062	0.0148	0.1215	0.1113	ethnic3	3
	0.944	0.9546	0.0112	0.0153	0.1235	2.0326	ethnic3	4

rho	true	mean	bias	mse	sdb	se	param	method
0.3	-0.472	-0.4498	0.0219	0.0036	0.0559	0.0511	sex	1
	-0.472	-0.4749	-0.0031	0.0030	0.0546	0.0538	sex	2
	-0.472	-0.4739	-0.0022	0.0032	0.0563	0.0514	sex	3
	-0.472	-0.4742	-0.0025	0.0032	0.0561	0.0539	sex	4
	-0.800	-0.7581	0.0419	0.0094	0.0874	0.0800	pov	1
	-0.800	-0.8070	-0.0070	0.0076	0.0870	0.0848	pov	2
	-0.800	-0.8050	-0.0050	0.0078	0.0883	0.0809	pov	3
	-0.800	-0.8105	-0.0105	0.0081	0.0894	0.0849	pov	4
	0.100	0.0891	-0.0109	0.0048	0.0685	0.0591	hw1	1
	0.100	0.0950	-0.0050	0.0045	0.0671	0.0621	hw1	2
	0.100	0.0954	-0.0046	0.0046	0.0675	0.0592	hw1	3
	0.100	0.0949	-0.0051	0.0046	0.0680	0.0621	hw1	4
	0.300	0.2770	-0.0230	0.0077	0.0848	0.0743	hw2	1
	0.300	0.2910	-0.0090	0.0068	0.0818	0.0781	hw2	2
	0.300	0.2921	-0.0079	0.0070	0.0832	0.0746	hw2	3
	0.300	0.2919	-0.0081	0.0067	0.0816	0.0782	hw2	4
	-2.359	-2.2272	0.1317	0.0271	0.0989	0.0943	ethnic1	1
	-2.359	-2.3573	0.0015	0.0114	0.1069	0.1010	ethnic1	2
	-2.359	-2.3549	0.0040	0.0114	0.1068	0.0965	ethnic1	3
	-2.359	-2.3633	-0.0044	0.0116	0.1078	0.1012	ethnic1	4
	-1.887	-1.7904	0.0967	0.0186	0.0962	0.0879	ethnic2	1
	-1.887	-1.8923	-0.0052	0.0099	0.0991	0.0928	ethnic2	2
	-1.887	-1.8885	-0.0014	0.0101	0.1003	0.0885	ethnic2	3
	-1.887	-1.8945	-0.0074	0.0101	0.1001	0.0929	ethnic2	4
	0.944	0.8954	-0.0481	0.0166	0.1196	0.1113	ethnic3	1
	0.944	0.9435	-0.0001	0.0150	0.1226	0.1169	ethnic3	2
	0.944	0.9399	-0.0037	0.0155	0.1244	0.1112	ethnic3	3
	0.944	0.9450	0.0015	0.0151	0.1231	0.1171	ethnic3	4

rho	true	mean	bias	mse	sdb	se	param	method
0.5	-0.472	-0.4547	0.0171	0.0036	0.0571	0.0532	sex	1
	-0.472	-0.4729	-0.0011	0.0033	0.0571	0.0550	sex	2
	-0.472	-0.4710	0.0008	0.0033	0.0576	0.0514	sex	3
	-0.472	-0.4720	-0.0002	0.0036	0.0597	0.0552	sex	4
	-0.800	-0.7770	0.0230	0.0082	0.0875	0.0833	pov	1
	-0.800	-0.8080	-0.0080	0.0078	0.0879	0.0865	pov	2
	-0.800	-0.8063	-0.0063	0.0081	0.0897	0.0809	pov	3
	-0.800	-0.8090	-0.0090	0.0081	0.0898	0.0867	pov	4
	0.100	0.0935	-0.0065	0.0042	0.0644	0.0613	hw1	1
	0.100	0.0968	-0.0032	0.0042	0.0648	0.0633	hw1	2
	0.100	0.0961	-0.0039	0.0042	0.0651	0.0592	hw1	3
	0.100	0.0990	-0.0010	0.0042	0.0649	0.0635	hw1	4
	0.300	0.2839	-0.0161	0.0073	0.0840	0.0774	hw2	1
	0.300	0.2951	-0.0049	0.0068	0.0820	0.0799	hw2	2
	0.300	0.2909	-0.0091	0.0064	0.0792	0.0746	hw2	3
	0.300	0.2958	-0.0042	0.0067	0.0820	0.0802	hw2	4
	-2.359	-2.2764	0.0825	0.0170	0.1012	0.0984	ethnic1	1
	-2.359	-2.3621	-0.0032	0.0110	0.1050	0.1034	ethnic1	2
	-2.359	-2.3592	-0.0003	0.0115	0.1074	0.0966	ethnic1	3
	-2.359	-2.3673	-0.0084	0.0113	0.1062	0.1039	ethnic1	4
	-1.887	-1.8258	0.0613	0.0122	0.0919	0.0910	ethnic2	1
	-1.887	-1.8902	-0.0031	0.0094	0.0970	0.0949	ethnic2	2
	-1.887	-1.8857	0.0014	0.0099	0.0997	0.0887	ethnic2	3
	-1.887	-1.8914	-0.0043	0.0091	0.0954	0.0952	ethnic2	4
	0.944	0.9118	-0.0317	0.0158	0.1217	0.1157	ethnic3	1
	0.944	0.9455	0.0020	0.0150	0.1223	0.1189	ethnic3	2
	0.944	0.9433	-0.0002	0.0154	0.1243	0.1109	ethnic3	3
	0.944	0.9447	0.0012	0.0155	0.1247	0.1192	ethnic3	4

rho	true	mean	bias	mse	sdb	se	param	method
0.7	-0.472	-0.4617	0.0100	0.0037	0.0597	0.0546	sex	1
	-0.472	-0.4745	-0.0027	0.0034	0.0583	0.0560	sex	2
.	-0.472	-0.4725	-0.0007	0.0034	0.0583	0.0513	sex	3
	-0.472	-0.4758	-0.0040	0.0034	0.0584	0.0565	sex	4
	-0.800	-0.7815	0.0185	0.0088	0.0918	0.0853	pov	1
	-0.800	-0.8017	-0.0017	0.0085	0.0919	0.0882	pov	2
	-0.800	-0.8010	-0.0010	0.0088	0.0935	0.0809	pov	3
	-0.800	-0.8034	-0.0034	0.0084	0.0915	0.0889	pov	4
	0.100	0.0891	-0.0109	0.0050	0.0701	0.0630	hw1	1
	0.100	0.0912	-0.0088	0.0048	0.0689	0.0646	hw1	2
	0.100	0.0940	-0.0060	0.0049	0.0699	0.0594	hw1	3
	0.100	0.0935	-0.0065	0.0047	0.0682	0.0652	hw1	4
	0.300	0.2908	-0.0092	0.0079	0.0884	0.0793	hw2	1
	0.300	0.2976	-0.0024	0.0072	0.0848	0.0812	hw2	2
	0.300	0.3015	0.0015	0.0077	0.0878	0.0743	hw2	3
	0.300	0.2988	-0.0012	0.0073	0.0852	0.0819	hw2	4
	-2.359	-2.3028	0.0561	0.0141	0.1049	0.1012	ethnic1	1
	-2.359	-2.3611	-0.0022	0.0113	0.1063	0.1053	ethnic1	2
	-2.359	-2.3581	0.0007	0.0110	0.1047	0.0966	ethnic1	3
	-2.359	-2.3694	-0.0105	0.0113	0.1057	0.1064	ethnic1	4
	-1.887	-1.8532	0.0339	0.0127	0.1073	0.0933	ethnic2	1
	-1.887	-1.8911	-0.0040	0.0114	0.1067	0.0965	ethnic2	2
	-1.887	-1.8895	-0.0024	0.0115	0.1073	0.0885	ethnic2	3
	-1.887	-1.8990	-0.0119	0.0114	0.1060	0.0976	ethnic2	4
	0.944	0.9212	-0.0224	0.0180	0.1323	0.1182	ethnic3	1
	0.944	0.9391	-0.0044	0.0173	0.1316	0.1218	ethnic3	2
	0.944	0.9364	-0.0072	0.0179	0.1335	0.1113	ethnic3	3
	0.944	0.9447	0.0012	0.0167	0.1292	0.1229	ethnic3	4

rho	true	mean	bias	mse	sdb	se	param	method
0.9	-0.472	-0.4678	0.0040	0.0043	0.0653	0.0557	sex	1
	-0.472	-0.4714	0.0004	0.0038	0.0616	0.0565	sex	2
	-0.472	-0.4719	-0.0001	0.0039	0.0621	0.0513	sex	3
	-0.472	-0.4756	-0.0039	0.0035	0.0592	0.0580	sex	4
	-0.800	-0.7941	0.0059	0.0096	0.0977	0.0878	pov	1
	-0.800	-0.8032	-0.0032	0.0089	0.0941	0.0891	pov	2
	-0.800	-0.8007	-0.0007	0.0082	0.0904	0.0808	pov	3
	-0.800	-0.8116	-0.0116	0.0087	0.0927	0.0912	pov	4
	0.100	0.0992	-0.0008	0.0054	0.0732	0.0644	hw1	1
	0.100	0.1009	0.0009	0.0051	0.0715	0.0652	hw1	2
	0.100	0.0978	-0.0022	0.0054	0.0736	0.0593	hw1	3
	0.100	0.1008	0.0008	0.0045	0.0672	0.0668	hw1	4
	0.300	0.3021	0.0021	0.0093	0.0966	0.0806	hw2	1
	0.300	0.3038	0.0038	0.0084	0.0917	0.0821	hw2	2
	0.300	0.3062	0.0062	0.0086	0.0928	0.0744	hw2	3
	0.300	0.3071	0.0071	0.0080	0.0889	0.0841	hw2	4
	-2.359	-2.3419	0.0169	0.0123	0.1098	0.1040	ethnic1	1
	-2.359	-2.3665	-0.0077	0.0120	0.1093	0.1064	ethnic1	2
	-2.359	-2.3621	-0.0032	0.0128	0.1129	0.0965	ethnic1	3
	-2.359	-2.3926	-0.0337	0.0135	0.1111	0.1096	ethnic1	4
	-1.887	-1.8661	0.0210	0.0121	0.1078	0.0955	ethnic2	1
	-1.887	-1.8882	-0.0011	0.0105	0.1024	0.0974	ethnic2	2
	-1.887	-1.8853	0.0019	0.0103	0.1013	0.0885	ethnic2	3
	-1.887	-1.9081	-0.0210	0.0106	0.1009	0.1004	ethnic2	4
	0.944	0.9417	-0.0019	0.0199	0.1410	0.1208	ethnic3	1
	0.944	0.9543	0.0108	0.0173	0.1312	0.1232	ethnic3	2
	0.944	0.9454	0.0018	0.0175	0.1322	0.1111	ethnic3	3
	0.944	0.9610	0.0175	0.0170	0.1293	0.1258	ethnic3	4

Table 3

NAEP reporting group estimates and SEs - all items, 2PL

	(3 level) 1 factor	(4 level) 1 factor	(3 level) 2 factor	(3 level) 3 factor
male	0			
femal	-.051 (.019)	-.050 (.019)	-.053 (.020)	-.052 (.020)
white	0			
black	-.533 (.031)	-.530 (.031)	-.551 (.033)	-.568 (.032)
hispa	-.395 (.029)	-.394 (.030)	-.414 (.030)	-.423 (.031)
as/pa	-.269 (.082)	-.269 (.083)	-.220 (.084)	-.274 (.086)
amind	-.458 (.062)	-.458 (.061)	-.469 (.065)	-.500 (.065)
other	-.447 (.404)	-.421 (.395)	-.678 (.488)	-.727 (.593)
NE	0			
SE	-.016 (.052)	-.004 (.061)	-.019 (.049)	-.038 (.051)
Cent	-.059 (.050)	-.054 (.064)	-.062 (.048)	-.043 (.047)
West	-.119 (.047)	-.109 (.056)	-.118 (.043)	-.108 (.055)
extru	0			
lomet	-.128 (.071)	-.140 (.079)	-.142 (.077)	-.202 (.076)
himet	.359 (.089)	.381 (.083)	.366 (.081)	.387 (.095)
manct	.172 (.065)	.123 (.079)	.138 (.073)	.102 (.075)
urbfr	.213 (.071)	.209 (.077)	.155 (.077)	.105 (.080)
medct	.111 (.061)	.081 (.071)	.068 (.071)	.076 (.070)
smplc	.043 (.062)	.026 (.067)	.009 (.068)	-.005 (.067)
nfnhs	0			
finhs	.136 (.050)	.135 (.050)	.230 (.053)	.206 (.055)
smcol	.404 (.059)	.406 (.059)	.448 (.061)	.446 (.064)
colgr	.428 (.047)	.425 (.047)	.481 (.049)	.468 (.051)
DK	.155 (.045)	.153 (.046)	.212 (.048)	.199 (.050)
nores	.059 (.124)	.021 (.125)	.211 (.129)	.160 (.121)
s^2_PSU	-	.013 (.006)	-	
s^2_sch	.061 (.006)	.054 (.007)	.056 (.006)	.057 (.007)
s_1^2	1.0	1.0	1.0	1.0
s_2^2			1.0	1.0
s_3^2				1.0
log Lmax	-103,013.51	-103,007.14	-102,219.51	-102,014.57

Table 4
NAEP reporting group estimates
and SEs - all items, MIMIC

	(3 level) 1 factor
male	0
femal	.013 (.033)
white	0
black	-.997 (.079)
hispa	-.812 (.072)
as/pa	-.468 (.156)
amind	-.748 (.121)
other	-.571 (.599)
NE	0
SE	-.043 (.089)
Cent	.033 (.087)
West	-.288 (.088)
extru	0
lomet	-.393 (.123)
himet	.742 (.118)
manct	.178 (.104)
urbfr	.283 (.112)
medct	.226 (.097)
smplc	.045 (.098)
nfnhs	0
finhs	.285 (.088)
smcol	.815 (.116)
colgr	.841 (.095)
DK	.344 (.080)
nores	-.139 (.171)
s^2_PSU	-
s^2_sch	.257 (.038)
s_1^2	1.945 (.238)
s_2^2	
s_3^2	
log Lmax	-102,562.07