Journal of Educational and Behavioral Statistics
Spring 1997, Vol. 22, No. 1, pp. 47–76

Multilevel Item Response Models: An Approach to Errors in Variables Regression

Raymond J. Adams
Australian Council for Educational Research
Mark Wilson
University of California, Berkeley
Margaret Wu
Australian Council for Educational Research

Keywords: *EM algorithm, item response models, measurement error, multilevel models*

In this article we show how certain analytic problems that arise when one attempts to use latent variables as outcomes in regression analyses can be addressed by taking a multilevel perspective on item response modeling. Under a multilevel, or hierarchical, perspective we cast the item response model as a within-student model and the student population distribution as a between-student model. Taking this perspective leads naturally to an extension of the student population model to include a range of student-level variables, and it invites the possibility of further extending the models to additional levels so that multilevel models can be applied with latent outcome variables. In the two-level case, the model that we employ is formally equivalent to the plausible value procedures that are used as part of the National Assessment of Educational Progress (NAEP), but we present the method for a different class of measurement models, and we use a simultaneous estimation method rather than two-step estimation. In our application of the models to the appropriate treatment of measurement error in the dependent variable of a between-student regression, we also illustrate the adequacy of some approximate procedures that are used in NAEP.

Over recent years the development of models for multilevel, or hierarchical, data structures has been an active area of methodological research (for recent summaries see Bock, 1989; Goldstein, 1987; Raudenbush, 1988). These models have developed from the recognition that in many research settings data are collected in a hierarchical form. School students are nested within schools, schools within school systems, and so on. Models that recognize these hierarchies have proven useful for solving the technical problems that

---

All of the authors were supported by NSF Grant RED-9255272. The first author was partially supported by a National Academy of Education Spencer Fellowship, and the second author was partially supported by Contract N00014-91-J-1523 from the Cognitive Sciences Program, Office of Naval Research.

arise when traditional approaches and models are applied to nested data, and they have led to an improved conceptualization of research questions in hierarchical settings (Raudenbush, 1988).

While it appears that the first and perhaps the motivating applications of hierarchical models in education were to situations where students are nested within classrooms, it is clear that the conceptualization of data structures as hierarchical can also be of value in other contexts: for modeling growth (Bryk & Raudenbush, 1987; Goldstein, 1987), in fitting random effects models in meta-analysis contexts (Raudenbush & Bryk, 1985), and for errors-in-variables regression analysis (Adams, 1989; Goldstein, 1987). Work on nonlinear multilevel models has also been done (Anderson & Aitkin, 1985; Goldstein, 1991; Stiratelli, Laird, & Waire, 1984; Wong & Mason, 1985), although this has primarily occurred outside of the educational research literature.

In psychometrics the recently developed structural measurement models and their accompanying marginal maximum likelihood estimation methods (Bock & Aitkin, 1981; Bock & Lieberman, 1970) are also nonlinear multilevel models. These models, which have been strongly advocated (e.g., Holland, 1990), require the assumption that individuals have been sampled from a population in addition to the specification of a model for generating item responses. The combination of these two models can be viewed as a nonlinear multilevel model. In the multilevel formulation the item response model can be viewed as a within-student model, while the population model can be viewed as a between-student model.

Although motivated and cast somewhat differently, structural measurement models of the type we discuss have been extensively developed and applied as part of the National Assessment of Educational Progress (NAEP) (Beaton, 1987; Mislevy, 1991; Zwick, 1992). In this article we illustrate the connection between hierarchical modeling and the structural measurement models as applied in NAEP by describing the structural measurement models as two-level nonlinear multilevel models. Due to the fact that we use a model slightly different from that used in NAEP and the previous relative inaccessibility of the detailed material on the NAEP estimation, we describe the model and its estimation in a little bit more detail than some may regard as necessary. We report some simulations to illustrate and test the estimation strategy. In the final section we use the model to fit errors in variables regression models to five real data sets where the independent variable is latent. For each of the five data sets we compare the results to those that would have been achieved if the error had been ignored and to those that would have been achieved by secondary analyses if they had used plausible values.

A Two-Level Formulation

The description of a structural item response model requires the specification of two components—a conditional item response model $f_{\mathbf{x}}(\mathbf{x};\ \boldsymbol{\xi}|\theta)$ and a population model $f_\theta(\theta; \boldsymbol{\alpha})$, where $\mathbf{x}$ is a vector of observations on items, $\boldsymbol{\xi}$ is a vector of parameters that describe those items, θ is a latent random variable (typically ability), and $\boldsymbol{\alpha}$ symbolizes a set of parameters that characterize the distribution of θ. The population model describes the between-student variation in the latent trait of interest, and the conditional item response model describes the probability of observing a set of item responses conditional upon the level of an individual on the latent trait of interest. In the case where the population is regarded as normal, $\boldsymbol{\alpha} = (\mu, \sigma^2)$. This is a structural model because θ is a random variable, that is, it does not have a fixed unknown value. When the θ are fixed, the model is a functional model and would be written $f_\mathbf{x}(\mathbf{x}; \boldsymbol{\xi}, \theta)$ (de Leeuw & Verhelst, 1986).

For a structural model the probability of the response vector $\mathbf{x}$ of a student randomly sampled from the population is

$$f_\mathbf{x}(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\alpha}) = \int_\theta f_\mathbf{x}(\mathbf{x}; \boldsymbol{\xi}|\theta) f_\theta(\theta; \boldsymbol{\alpha})\, d\theta, \qquad (1)$$

and it follows that the likelihood is

$$\Lambda = \prod_{n=1}^{N} f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}, \boldsymbol{\alpha}), \qquad (2)$$

where N is the total number of sampled students.

Note, however, that $f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta)$ can be considered as a within-student model. It is the model that describes the responses of student n. If μ is the mean of the population distribution, f_θ, then we can write

$$\theta_n = \mu + E_n, \qquad (3)$$

where the distribution of E_n is the same as that for θ_n, but translated to have a mean of zero.

Equation 3 is a very simple linear model for variation between students in θ. Hence we identify the structural model as a two-level nonlinear hierarchical model. In this case the population is undifferentiated, in the sense that every individual is regarded as sampled from an identical population distribution.

A natural extension of (3) is to replace the mean, μ, with the regression model $\mathbf{Y}'_n\boldsymbol{\beta}$, where $\mathbf{Y}_n$ is a vector of u, fixed and known values for student n, and $\boldsymbol{\beta}$ is the corresponding vector of regression coefficients. For example, $\mathbf{Y}_n$ could be constituted of student variables such as gender, socioeconomic status, and major. Then the population model becomes

$$\theta_n = \mathbf{Y}'_n\boldsymbol{\beta} + E_n, \qquad (4)$$

where we assume that the E_n are independently and identically normally distributed with mean zero and variance σ^2 so that (4) is equivalent to

$$f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})'(\theta_n - \mathbf{Y}_n'\boldsymbol{\beta}) \right], \quad (5)$$

a normal with mean $\mathbf{Y}_n'\boldsymbol{\beta}$ and variance σ^2. If (5) is used as the population model, then the parameters to be estimated are $\boldsymbol{\beta}$, σ^2, and $\boldsymbol{\xi}$.

Both (3) and (4) can be generalized by allowing alternative distributions for E_n. For example, a step distribution that is specified by a set of nodes and weights (or densities) for each node was suggested by Tjur (1982) and Cressie and Holland (1983) and is implemented in BILOG (Mislevy & Bock, 1983) for (3) but not (4). In this article we will, however, restrict our attention to the normal case.

We see four potential advantages in extending the two-level structural model to the model with differentiated populations as given in (4). First, direct estimation of the population parameters (for example $\boldsymbol{\beta}$, σ^2) from the item responses obviates the problem of the bias introduced by two-step estimation (Mislevy, 1984), which typically proceeds by estimating person abilities and then using them in subsequent analyses under the assumptions of independence and identically distributed error terms. It is this motivation that has driven the development of models of this type for use in NAEP (Mislevy, Beaton, Kaplan, & Sheehan, 1992). Second, the use of student-level variables, $\mathbf{Y}_n$, can lead to increased precision in estimation of the item parameters, $\boldsymbol{\xi}$ (Mislevy, 1987; Mislevy & Sheehan, 1989a). Third, use of the same information can lead to increased precision in the estimation of person (ability) parameters, θ_n. That is, when individual abilities are estimated as the expected values of the marginal posterior, they will have a smaller mean squared error when the collateral information is employed. However, employing collateral information in this way is not without certain difficulties: When collateral information is used, point estimates of ability are biased, and, as Mislevy (1987) has stated, they are not "unequivocally 'better' for all applications" (p. 90). Finally, Mislevy and Sheehan (1989b) have shown that to ensure consistent item parameter estimates, collateral information must be used in estimation where it was used in item selection. For example, when students at different grade levels are given different test forms, matched to their expected abilities through age or grade information, subsequent marginal maximum likelihood (MML) parameter estimation will be inconsistent unless the age or grade level information is employed as a collateral variable.

The Item Response Model

For the item response model we will restrict our attention to the random coefficients multinominal logit (RCML) recently described by Adams and Wilson (1996). We use the RCML because it is a very general form of the

Rasch model encompassing the simple logistic model (Rasch, 1980), the rating scale model (Andrich, 1978), the partial credit model (Masters, 1982), FACETS (Linacre, 1989), linear logistic models (Fischer, 1983), and the ordered partition model (Wilson, 1992; Wilson & Adams, 1993) and may be used to develop many others. In addition to this generality, the RCML inherits the fundamental measurement properties of the Rasch family. Mislevy (1985) presents a similar procedure for the three-parameter logistic model, and it is also possible to use Muraki's (1992) generalized partial credit model with a joint estimation algorithm like the one we detail below.

To describe the RCML we suppose we have I items indexed $i = 1, \ldots, I$, and each item admits $K_i + 1$ response alternatives indexed $k = 0, 1, \ldots, K_i$. We then use the vector valued random variable $\mathbf{X}_i$ to indicate the $K_i + 1$ possible responses to item i. That is, $\mathbf{X}_i' = (X_{i1}, X_{i2}, \ldots, X_{iK_i})$, where

$$X_{ij} = \begin{cases} 1 & \text{if response to item } i \text{ is in category } j \\ 0 & \text{otherwise} \end{cases}.$$

A response in category zero is denoted by a vector of zeros. This effectively makes the zero category a reference category and is necessary for model identification. The choice of this as the reference category is arbitrary and does not affect the generality of the model. We collect the $\mathbf{X}_i$ together into a single vector $\mathbf{X}' = (\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_I')$, which is a vector valued random variable that is called a response pattern. Particular instances of each of these random variables are indicated by their lowercase equivalents: $\mathbf{x}$, $\mathbf{x}_i$, and x_{ik}.

The items are described by p difficulty parameters which are given by the vector $\boldsymbol{\xi}' = (\xi_1, \xi_2, \ldots, \xi_p)$. Linear combinations of these are used in the response probability model to describe the empirical characteristics of the response categories of each item. These linear combinations are defined by the design vectors $\mathbf{a}_{jk}$ ($j = 1, \ldots, I$; $k = 1, \ldots, K_i$), which can be denoted by the design matrix $\mathbf{A}' = (\mathbf{a}_{11}, \mathbf{a}_{12}, \ldots, \mathbf{a}_{1K_1}, \mathbf{a}_{21}, \ldots, \mathbf{a}_{2K_2}, \ldots, \mathbf{a}_{I1}, \ldots, \mathbf{a}_{IK_I})$. This approach to imposing a linear model on the item parameters allows us to write a general model that includes the wide class of existing Rasch models mentioned above and to develop new types of Rasch models, for example, the item bundles models of Wilson and Adams (1995).

An additional feature of the RCML is the introduction of a scoring function which allows the description of the score, or *performance level*, that is assigned to each response type. To do this we introduce the notion of a response score b_{ij} which gives the performance level of an observed response in category j of item i. The b_{ij} can be collected in a vector as $\mathbf{b}' = (b_{11}, b_{12}, \ldots, b_{1K_1}, b_{21}, b_{22}, \ldots, b_{2K_2}, \ldots, b_{I1}, \ldots, b_{IK_I})$. (By definition, the score for a response in the zero category is zero, but other responses may also be scored zero).

In the majority of Rasch model formulations there has been a one-to-one matching between the category to which a response belongs and the score that is allocated to the observation. In the simple logistic model, for example, it has been standard practice to use the labels "0" and "1" to indicate both the categories of performance and the scores. A similar practice has been followed with the rating scale and partial credit models, where each category of performance is seen as indicating a different level of performance. The use of **b** as a scoring function allows a more flexible relationship between the quality of a response and the level of performance that it reflects. Examples of where this is applicable are given in Kelderman (1989) and Wilson (1992).

We now write the RCML item response probability model as

$$\Pr(X_{ij} = 1 | \theta) = \frac{\exp(b_{ij}\theta + \mathbf{a}'_{ij}\boldsymbol{\xi})}{\sum_{k=1}^{K_i} \exp(b_{ik}\theta + \mathbf{a}'_{ik}\boldsymbol{\xi})},$$

and a response vector probability model as

$$\begin{aligned}\Pr(\mathbf{X} = \mathbf{x} | \theta) &= \Psi(\theta, \boldsymbol{\xi}) \exp\{\mathbf{x}'(\mathbf{b}\theta + \mathbf{A}\boldsymbol{\xi})\} \\ &= f_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\xi} | \theta), \end{aligned} \tag{6}$$

$$\text{with } \Psi(\theta, \boldsymbol{\xi}) = \left[\sum_{\mathbf{z} \in \boldsymbol{\Omega}} \exp\{\mathbf{z}'(\mathbf{b}\theta + \mathbf{A}\boldsymbol{\xi})\}\right]^{-1},$$

and where $\boldsymbol{\Omega}$ is the set of all possible response vectors.

Estimating the Model

In this section we show how either a Newton-Raphson method or the EM algorithm of Dempster, Laird, and Rubin (1977) can be used to jointly produce maximum likelihood estimates of the item and population parameters. By jointly estimating the item parameters and population parameters, our approach differs slightly from current practice in NAEP (Johnson, Mazzeo, & Kline, 1993; Thomas, 1992). To ease the computational burden, NAEP uses a two-step approach where the parameters are first estimated without the use of conditioning variables and in a second phase the item parameters are fixed at their estimated values while the population parameters are estimated as an intermediate step in the generation of plausible values.[1] In our case, the use of Rasch-type models and the restriction to one dimension make it easier to jointly estimate the items and population parameters.

If we observe data from a sample of N students and denote all of the observed data (that is, item responses and student-level data) as $\mathbf{X}$, then the likelihood induced by (1) is

$$\Lambda(\boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2 | \mathbf{X}) = \prod_{n=1}^{N} \int_{\theta_n} f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi} | \theta_n) f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)\, d\theta_n, \tag{7}$$

and the log likelihood is

$$\lambda(\boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2 | \mathbf{X}) = \sum_{n=1}^{N} \log \int_{\theta_n} f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi} | \theta_n) f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)\, d\theta_n. \tag{8}$$

Differentiating with respect to each of the parameters and defining the marginal posterior as

$$h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2 | \mathbf{x}_n) = \frac{f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi} | \theta_n) f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{f_{\mathbf{x}}(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)} \tag{9}$$

provides the following likelihood equations:

$$\begin{aligned}
\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\xi}} &= \frac{\partial}{\partial \boldsymbol{\xi}} \left\{ \sum_{n=1}^{N} \log \int_{\theta_n} f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi} | \theta_n) f_\theta(\theta_n | \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)\, d\theta_n \right\} \\
&= \sum_{n=1}^{N} \frac{1}{f_{\mathbf{x}}(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)} \int_{\theta_n} \frac{\partial f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi} | \theta_n)}{\partial \boldsymbol{\xi}} f_\theta(\theta_n | \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)\, d\theta_n \\
&= \sum_{n=1}^{N} \int_{\theta_n} \frac{\partial \log f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi} | \theta_n)}{\partial \boldsymbol{\xi}} h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2 | \mathbf{x}_n)\, d\theta_n \\
&= \sum_{n=1}^{N} \int_{\theta_n} \frac{\partial}{\partial \boldsymbol{\xi}} \left[\mathbf{x}_n'(\mathbf{b}\theta_n + \mathbf{A}\boldsymbol{\xi}) - \log \sum_{\mathbf{z} \in \boldsymbol{\Omega}} \exp \mathbf{z}'(\mathbf{b}\theta_n + \mathbf{A}\boldsymbol{\xi}) \right] \\
&\quad \times h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2 | \mathbf{x}_n)\, d\theta_n \\
&= \sum_{n=1}^{N} \int_{\theta_n} \left[\mathbf{A}'\mathbf{x}_n - \mathbf{A}'\boldsymbol{\Psi}(\theta_n, \boldsymbol{\xi}) \sum_{\mathbf{z} \in \boldsymbol{\Omega}} \mathbf{z} \exp \mathbf{z}'(\mathbf{b}\theta_n + \mathbf{A}\boldsymbol{\xi}) \right] \\
&\quad \times h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2 | \mathbf{x}_n)\, d\theta_n \\
&= \mathbf{A}' \sum_{n=1}^{N} \left[\mathbf{x}_n - \int_{\theta_n} E_{\mathbf{z}}(\mathbf{z} | \theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2 | \mathbf{x}_n)\, d\theta_n \right] \\
&= \mathbf{0},
\end{aligned} \tag{10}$$

where $E_{\mathbf{z}}(\mathbf{z}|\theta_n) = \boldsymbol{\Psi}(\theta_n, \boldsymbol{\xi}) \sum_{\mathbf{z}\in\boldsymbol{\Omega}} \mathbf{z} \exp\{\mathbf{z}'(\mathbf{b}\theta_n + \mathbf{A}\boldsymbol{\xi})\}$;

$$\begin{aligned}
\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\beta}} &= \sum_{n=1}^{N} \int_{\theta_n} \frac{\partial \log f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial\boldsymbol{\beta}} h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \\
&= \sigma^{-2} \sum_{n=1}^{N} \int_{\theta_n} (\mathbf{Y}_n\theta_n - \mathbf{Y}_n\mathbf{Y}_n'\boldsymbol{\beta})\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \\
&= \sigma^{-2} \sum_{n=1}^{N} \left\{\int_{\theta_n} \mathbf{Y}_n\theta_n\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n - \mathbf{Y}_n\mathbf{Y}_n'\boldsymbol{\beta}\right\} \\
&= \sigma^{-2} \sum_{n=1}^{N} \left\{\int_{\theta_n} \mathbf{Y}_n\theta_n\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\right\} \\
&\quad - \sigma^{-2} \sum_{n=1}^{N} \mathbf{Y}_n\mathbf{Y}_n'\boldsymbol{\beta} \\
&= \sigma^{-2}\left\{\sum_{n=1}^{N} \mathbf{Y}_n\bar{\theta}_n - \sum_{n=1}^{N} \mathbf{Y}_n\mathbf{Y}_n'\boldsymbol{\beta}\right\} \\
&= \mathbf{0}, \qquad (11)
\end{aligned}$$

where $\bar{\theta}_n = \int_{\theta_n} \theta_n h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$; and

$$\begin{aligned}
\frac{\partial\boldsymbol{\lambda}}{\partial\sigma^2} &= \sum_{n=1}^{N} \int_{\theta_n} \frac{\partial \log f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial\sigma^2} h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \\
&= \sum_{n=1}^{N} \int_{\theta_n} -\frac{1}{2\sigma^4} [\sigma^2 - (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2]\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \\
&= -\frac{N}{2\sigma^4}\left[\sigma^2 - \frac{1}{N}\sum_{n=1}^{N} \int_{\theta_n} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\right] \\
&= 0. \qquad (12)
\end{aligned}$$

To jointly solve these likelihood equations, we have considered a Newton-Raphson algorithm and an EM algorithm. In our implementations, both algorithms require the hessian matrix, the derivation of which we have presented in the Appendix. The Newton-Raphson algorithm requires the calculation and inversion of the expected hessian at each iteration, but with the EM algorithm the hessian is calculated and inverted only at the solution to provide asymptotic standard errors for the parameter estimates.

For the Newton-Raphson algorithm, let $\boldsymbol{\zeta}' = (\boldsymbol{\xi}', \boldsymbol{\beta}', \sigma^2)$ be the vector containing all of the parameters, and then define $\boldsymbol{\zeta}^{(i)}$ as the parameter estimates after iteration i, $\boldsymbol{\zeta}^{(0)}$ as initial estimates of the parameters, $\mathbf{S}^{(i)}$ as the scores vector evaluated at $\boldsymbol{\zeta}^{(i)}$, and $\mathbf{E}^{(i)}$ as the hessian evaluated at $\boldsymbol{\zeta}^{(i)}$. Then the Newton equation,

$$\boldsymbol{\zeta}^{(i+1)} = \boldsymbol{\zeta}^{(i)} + \mathbf{S}^{(i)}\mathbf{E}^{(i)^{-1}},$$

is used to iterate to a solution. In practice we have found this method to be somewhat unstable, and when it does converge each iteration can be quite time consuming.

Our current preferred strategy is to use the EM algorithm beginning with initial estimates $\boldsymbol{\zeta}^{(0)}$ and using (9) to calculate a provisional marginal posterior. This is the E-step of the algorithm. The likelihood equations (10), (11), and (12) are then solved, treating the posterior as known to produce updated parameter estimates. This is the M-step of the algorithm. Expressions equivalent to these are reported in Mislevy (1984, 1985) and Thomas (1992).

Because the integrals in the likelihood equations cannot be analytically simplified, we approximate them as follows. Let $\Theta_1, \Theta_2, \ldots, \Theta_Q$ be a set of fixed grid points with a constant difference between them, so that $\Theta_{i+1} - \Theta_i = \Delta_\theta$. Then the integral in likelihood equation (10) can be approximated using

$$\sum_{n=1}^{N}\left(\int_{\theta_n} E_{\mathbf{z}}(\mathbf{z}|\theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\right)$$

$$\approx \sum_{n=1}^{N}\sum_{q=1}^{Q} E_{\mathbf{z}}(\mathbf{z}|\Theta_q)\, h_\theta(\Theta_q; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, \Delta_\theta$$

and solved using a Newton-Raphson routine.

Similar approximations to the mean and second moments of the posterior are given by

$$\begin{aligned}\bar{\theta}_n &= \int_{\theta_n} \theta_n\, h_\theta(\theta_n, \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \\ &\approx \sum_{q=1}^{Q} \Theta_q h_\theta(\Theta_q, \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\Delta_\theta \\ &= \tilde{\theta}_n\end{aligned}$$

and

$$\int_{\theta_n} \theta_n^2\, h_\theta(\theta_n, \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \approx \sum_{q=1}^{Q} \Theta_q^2 h_\theta(\Theta_q, \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\Delta_\theta,$$

yielding

$$\hat{\boldsymbol{\beta}} = \left(\sum_{n=1}^{N} \mathbf{Y}_n\mathbf{Y}_n'\right)^{-1}\left(\sum_{n=1}^{N} \tilde{\theta}_n\mathbf{Y}_n\right)$$

and

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}\sum_{q=1}^{Q} (\Theta_q - \mathbf{Y}_n'\boldsymbol{\beta})^2\, h_\theta(\Theta_q; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\Delta_\theta$$

as computational formulas for (11) and (12). Similar approximations are used in the calculation of the hessian. After using an EM algorithm to compute maximum likelihood estimates for $\boldsymbol{\xi}$, $\boldsymbol{\beta}$, and σ^2, we calculate the negative inverse of the hessian to provide an asymptotic variance-covariance matrix for the parameter estimates.

Parameter Recovery

Simulation 1

To verify our approach to estimation, we have undertaken some exploratory simulations. In each of our simulations we have used 100 replications. While somewhat arbitrary, our choice of 100 replications was judged to be sufficient to provide us some basic empirical information with regard to the properties of the estimators, and the algorithm as we have implemented it.

In the first simulation we considered the case of a sample of 500 students responding to a partial credit test of five items, three of which have three response categories and two of which have four response categories. A single predictor variable Y that was correlated 0.90 with the latent ability variable θ was assumed. This configuration is unusual in that it corresponds to a case where the magnitude of the error in the outcome variable is large and the correlation between the latent ability and another known variable is very informative about the latent ability. As an example, it provides an extreme illustration of the difference between a two-step approach to estimation (i.e., an approach in which a measurement model is first fitted ignoring the collateral variable Y and the regression of θ on Y is then undertaken) and a one-step approach (i.e., joint estimation of the measurement and regression models).

Each simulated data set contained 500 simulated students. To simulate a latent ability θ_n and a collateral variable Y_n for each student, $n = 1, \ldots,$

500, we began by randomly generating a pair of independent unit normal deviates, s_n and t_n. Applying the linear transformations

$$\theta_n = \sqrt{.95}s_n + \sqrt{.05}t_n$$

and

$$Y_n = \sqrt{.95}s_n - \sqrt{.05}t_n,$$

we obtained θ and Y values with the desired correlation of 0.90. Item responses for each student were generated by assuming the RCML and the generated latent ability values for each student and the set of item parameters as given in Table 1. For the simulated values of θ and Y, it is easy to show that the generating model is

$$\theta_n = \beta_0 + \beta_1 Y_n + e_n,$$

TABLE 1
Generating values, means of recovered values for Simulation 1

Parameter	Generating value	Mean of estimates	Difference	Sampling variance	Mean of estimated error variance	Ratio of sampling variance to estimated variance
$\xi_1 \equiv \delta_{11}$	1.226	1.212	−0.014	0.022	0.022	0.986
$\xi_2 \equiv \delta_{12}$	0.595	0.597	0.002	0.012	0.012	0.989
$\xi_3 \equiv \delta_{21}$	0.998	0.992	−0.006	0.018	0.017	1.029
$\xi_4 \equiv \delta_{22}$	−0.103	−0.106	−0.003	0.012	0.013	0.935
$\xi_5 \equiv \delta_{23}$	−2.966	−2.987	−0.021	0.066	0.052	1.271
$\xi_6 \equiv \delta_{31}$	1.575	1.588	0.013	0.017	0.016	1.078
$\xi_7 \equiv \delta_{32}$	−1.828	−1.844	−0.016	0.025	0.023	1.108
$\xi_7 \equiv \delta_{33}$	−1.017	−0.999	0.018	0.049	0.040	1.220
$\xi_9 \equiv \delta_{41}$	1.168	1.176	0.008	0.021	0.014	1.447
$\xi_{10} \equiv \delta_{42}$	−1.177	−1.198	−0.021	0.012	0.014	0.838
$\xi_{11} \equiv \delta_{51}$	0.093	0.103	0.010	0.017	0.024	0.716
$\xi_{12} \equiv \delta_{52}$	0.432	0.434	0.002	0.026	0.026	0.987
β_0	0.000	−0.004	−0.004	0.001	0.001	0.671
β_1	0.900	0.904	0.004	0.002	0.002	1.316
σ^2	0.190	0.199	0.009	0.002	0.002	1.279

Note. Hotelling's T^2 for bias is 1.10 on 15, 85 *df* (n.s.).

where $\beta_0 = 0$, $\beta_1 = 0.9$, and $e_n \overset{\text{iid}}{\sim} N(0, 0.19)$. Furthermore, the model

$$\theta_n = \alpha_0 + \varepsilon_n,$$

where $\alpha_0 = 0$ and $\varepsilon_n \overset{\text{iid}}{\sim} N(0, 1.0)$, also holds.

We call the first of the above models the conditional model and the second the unconditional model. In each of the 100 replications both the conditional and unconditional models were fitted to the data. Expected a posteriori (EAP) predictions of ability were made from both models, and the unconditional EAP predictions were regressed onto Y using ordinary least squares.

In Table 1 we present the generating values for the parameters in the conditional model, the mean of the recovered values, the difference between the generating value and the mean of the recovered values, the between-replication variance in the parameter estimates, the mean of the asymptotic estimates of the parameter estimate variance, and the ratio of the sampling variance to the asymptotic estimates of the error variance. The labeling of the parameters with δ_{ij} is given to make the connection between the ξ parameters and the usual partial credit model notation (Masters, 1982).

The result of a Hotelling's T^2 test, which tests the hypothesis that the expected differences between generating and estimated values are all zero, is reported at the bottom of the table. This F value is not significant, and we accept a hypothesis of no bias in the parameter estimates. The results do, however, suggest some caution with regard to the use of the asymptotic approximation to the error variances. The range of the variance ratios is 0.671 to 1.447; this corresponds to asymptotic approximations to the standard error that vary from underestimation by about 20% to overestimation by about 20%.

Theoretical work (e.g., Mislevy, 1987; Mislevy & Sheehan, 1989a) has shown that the use of collateral information will lead to smaller error in the estimation of item parameters and reduce the mean squared error on ability prediction. The first panel of Table 2 gives the between-replication variance in item parameter estimates for the unconditional and conditional models and their ratio. The second panel reports the mean squared error in the respective EAP ability predictions. The results reported in Table 2 show that the extra information provided by the collateral variable Y has a negligible effect on the accuracy of item parameter estimates, but does lead to a substantial decrease in the mean squared error in ability predictions. In this simulation, the collateral variable is strongly related to the outcome θ and can therefore be expected to lead to improved ability prediction. In subsequent analyses, we note the magnitude of the improvement when the collateral information is less informative.

The negligible differences between the item parameter estimates for these analyses with and without conditioning variables are consistent with and supportive of the current NAEP practice of estimating the items parameters

TABLE 2
Sampling variation in item parameter estimates for conditional and unconditional models in Simulation 1

Parameter	Generating value	Sampling variance from the unconditional model	Sampling variance from the conditional model	Ratio
$\xi_1 \equiv \delta_{11}$	1.226	0.022	0.022	1.008
$\xi_2 \equiv \delta_{12}$	0.595	0.012	0.012	1.008
$\xi_3 \equiv \delta_{21}$	0.998	0.018	0.017	1.011
$\xi_4 \equiv \delta_{22}$	−0.103	0.013	0.013	1.018
$\xi_5 \equiv \delta_{23}$	−2.966	0.065	0.052	0.983
$\xi_6 \equiv \delta_{31}$	1.575	0.017	0.016	1.005
$\xi_7 \equiv \delta_{32}$	−1.828	0.025	0.023	0.989
$\xi_7 \equiv \delta_{33}$	−1.017	0.049	0.040	1.010
$\xi_9 \equiv \delta_{41}$	1.168	0.021	0.014	1.013
$\xi_{10} \equiv \delta_{42}$	−1.177	0.012	0.014	0.976
$\xi_{11} \equiv \delta_{51}$	0.093	0.017	0.024	0.988
$\xi_{12} \equiv \delta_{52}$	0.432	0.025	0.026	0.991

Note. Unconditional ability prediction mean squared error = 0.295. Conditional ability prediction mean squared error = 0.133.

without the use of conditioning variables, excepting those necessary to ensure consistency.

Finally, for this simulation we noted that the mean estimates of the regression coefficients that result from regressing unconditional predicted values of θ on Y using ordinary least squares are −0.001 and 0.637, and the mean R^2 is 0.573. This compares with the mean of the one-step conditional model estimates for the regression coefficients of −0.004 and 0.904 and a mean R^2 of 0.800 (compared to a generating R^2 of 0.810). The difference between the slope parameters highlights the attenuation due to measurement error that occurs in ordinary least squares two-step analyses.

Simulation 2

As the basis for our second simulation, we took the results of the analysis of a nine-item partial credit test of student understanding of science concepts related to the Earth and its place in the solar system. The development and use of the test is described in Adams, Doig, and Rosier (1991). For the population model, we used

$$\theta_n = \beta_0 + \beta_1 GRADE_n + \beta_2 SEX_n + \beta_3 (GRADE*SEX)_n + \beta_4 SES_n + e_n,$$

where $e_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$.

The generating values of the item parameters and the regression coefficients for the simulation are reported in Table 3 and were the product of fitting the model to the real data set. In simulating θ, the student-level variables, **Y**, in the actual data set and the generating regression parameters were used. The sample size for the simulations was 993—the number of students that were in the real data set for which we had a set of **Y** values.

As was described for the first simulation, each replication of the simulation included fitting both a conditional and an unconditional model and an ordinary least squares regression of the unconditional estimates of θ onto the collateral information. Additionally, we used an imputation-based approach paralleling the NAEP approach. Detailed descriptions of the NAEP methodology are provided elsewhere (e.g., Beaton, 1987; Johnson et al., 1993; Zwick, 1992) and will not be reported here. It is worth noting, however, that in our implementation we jointly estimated the item and population parameters before drawing the imputations (plausible values), and in evaluating the posterior we used a quadrature approach rather than the direct normal approximation or an approach with asymptotic corrections (Thomas, 1992). This was feasible because, in contrast to NAEP, we were dealing with a single latent ability distribution.

In Table 3 we present the generating values for the parameters in the model, the mean of the recovered values, the difference between the generating values and the mean of the recovered values, the between-replication variance in the parameter estimates, the mean of the asymptotic estimates of the parameter estimate variance, and the ratio of the sampling variance to the asymptotic estimates of the error variance.

As was the case in the previous simulation, the Hotelling's T^2 test is not significant, which suggests unbiased estimators. As we noted for the first simulation, the variation in the ratios of the sampling variance to the asymptotic estimates of the error variance suggests that some caution will be needed in the use of asymptotic estimates of the standard errors.

In Table 4 we compare the sampling variance of unconditional and conditional estimates of the item parameters and the mean squared errors of the ability predictions from the conditional and unconditional models. We note that the conditional estimates have a slightly smaller sampling variance than the unconditional estimates, and the mean squared error of the ability predictions is smaller for the conditional than the unconditional. The improvement, from the unconditional model to the conditional model, in the mean squared error of ability predictions is smaller for this simulation than it was for the previous simulation. This is because the conditioning variables are less strongly related to the outcome variable and because the test is four items longer.

Finally, for this simulation we are able to report three different estimators of the regression coefficients and R^2: estimates of the regression coefficients that result from regressing unconditional predicted values of θ on **Y** using

TABLE 3
Generating values, means of recovered values for Simulation 2

Parameter	Generating value	Mean of estimates	Difference	Sampling variance	Mean of estimated error variance	Ratio of sampling variance to estimated variance
$\xi_1 \equiv \delta_{11}$	1.219	1.233	0.014	0.011	0.008	1.342
$\xi_2 \equiv \delta_{12}$	0.597	0.603	0.006	0.005	0.006	0.898
$\xi_3 \equiv \delta_{21}$	0.999	1.008	0.009	0.007	0.007	0.991
$\xi_4 \equiv \delta_{22}$	−0.103	−0.092	0.011	0.007	0.007	0.996
$\xi_5 \equiv \delta_{23}$	−2.965	−3.033	−0.068	0.071	0.066	1.081
$\xi_6 \equiv \delta_{31}$	1.576	1.586	0.010	0.006	0.006	0.920
$\xi_7 \equiv \delta_{32}$	−1.828	−1.834	−0.006	0.016	0.015	1.039
$\xi_8 \equiv \delta_{33}$	−1.015	−1.023	−0.008	0.053	0.040	1.306
$\xi_9 \equiv \delta_{41}$	1.168	1.185	0.017	0.006	0.006	1.097
$\xi_{10} \equiv \delta_{42}$	−1.177	−1.178	−0.001	0.010	0.009	1.013
$\xi_{11} \equiv \delta_{51}$	0.094	0.090	−0.004	0.008	0.009	0.853
$\xi_{12} \equiv \delta_{52}$	0.431	0.425	−0.006	0.011	0.012	0.933
$\xi_{13} \equiv \delta_{53}$	0.409	0.417	0.008	0.010	0.010	0.974
$\xi_{14} \equiv \delta_{61}$	0.860	0.862	0.002	0.006	0.006	1.089
$\xi_{15} \equiv \delta_{62}$	−0.834	−0.835	−0.001	0.009	0.009	1.087
$\xi_{16} \equiv \delta_{71}$	0.084	0.086	0.002	0.012	0.010	1.244
$\xi_{17} \equiv \delta_{72}$	0.277	0.279	0.002	0.014	0.014	1.008
$\xi_{18} \equiv \delta_{73}$	0.940	0.947	0.007	0.010	0.011	0.880
$\xi_{19} \equiv \delta_{81}$	−0.669	−0.661	0.008	0.017	0.019	0.923
$\xi_{20} \equiv \delta_{82}$	−0.674	−0.678	−0.004	0.091	0.071	1.290
$\xi_{21} \equiv \delta_{83}$	3.673	3.685	0.012	0.065	0.058	1.130
$\xi_{22} \equiv \delta_{91}$	0.443	0.448	0.005	0.005	0.005	0.911
$\xi_{23} \equiv \delta_{92}$	−0.645	−0.657	−0.012	0.010	0.009	1.100
$\xi_{24} \equiv \delta_{93}$	−2.784	−2.795	−0.011	0.077	0.097	0.793
β_0	−0.780	−0.782	−0.002	0.002	0.002	1.084
β_1	0.657	0.655	−0.001	0.005	0.005	1.137
β_2	−0.002	−0.006	−0.004	0.004	0.004	1.103
β_3	0.136	0.135	−0.001	0.012	0.009	1.297
β_4	0.378	0.383	0.006	0.001	0.001	1.131
σ^2	0.388	0.389	0.001	0.001	0.001	1.067

Note. Hotelling's T^2 for bias is 0.81 on 30, 70 *df* (n.s.).

ordinary least squares, estimates that come from the conditional two-level model, and estimates that are derived from plausible values.

Table 5 shows that both the conditional two-level estimates and the plausible value estimates are close to the generating values, while, as expected, the OLS estimates underestimate the R^2 and the magnitude of the main effects β_1, β_3, and β_4. These two simulations suggest that the EM algorithm that we

TABLE 4
Sampling variation in item parameter estimates for conditional and unconditional models in Simulation 2

Parameter	Generating value	Sampling variance from the unconditional model	Sampling variance from the conditional model	Ratio
$\xi_1 \equiv \delta_{11}$	1.219	0.011	0.011	1.014
$\xi_2 \equiv \delta_{12}$	0.597	0.005	0.005	1.008
$\xi_3 \equiv \delta_{21}$	0.999	0.007	0.007	1.008
$\xi_4 \equiv \delta_{22}$	−0.103	0.007	0.007	1.010
$\xi_5 \equiv \delta_{23}$	−2.965	0.072	0.071	1.015
$\xi_6 \equiv \delta_{31}$	1.576	0.006	0.006	1.016
$\xi_7 \equiv \delta_{32}$	−1.828	0.016	0.016	1.010
$\xi_8 \equiv \delta_{33}$	−1.015	0.053	0.053	1.014
$\xi_9 \equiv \delta_{41}$	1.168	0.006	0.006	1.013
$\xi_{10} \equiv \delta_{42}$	−1.177	0.010	0.010	1.009
$\xi_{11} \equiv \delta_{51}$	0.094	0.008	0.008	1.017
$\xi_{12} \equiv \delta_{52}$	0.431	0.011	0.011	1.003
$\xi_{13} \equiv \delta_{53}$	0.409	0.010	0.010	1.010
$\xi_{14} \equiv \delta_{61}$	0.860	0.006	0.006	1.013
$\xi_{15} \equiv \delta_{62}$	−0.834	0.009	0.009	1.011
$\xi_{16} \equiv \delta_{71}$	0.084	0.012	0.012	1.013
$\xi_{17} \equiv \delta_{72}$	0.277	0.014	0.014	1.009
$\xi_{18} \equiv \delta_{73}$	0.940	0.010	0.010	1.007
$\xi_{19} \equiv \delta_{81}$	−0.669	0.017	0.017	1.010
$\xi_{20} \equiv \delta_{82}$	−0.674	0.092	0.091	1.009
$\xi_{21} \equiv \delta_{83}$	3.673	0.066	0.065	1.011
$\xi_{22} \equiv \delta_{91}$	0.443	0.005	0.005	1.014
$\xi_{23} \equiv \delta_{92}$	−0.645	0.010	0.010	1.008
$\xi_{24} \equiv \delta_{93}$	−2.784	0.078	0.077	1.014

Note. Unconditional ability prediction mean squared error = 0.134. Conditional ability prediction mean squared error = 0.122.

have implemented produces parameter estimates that are essentially unbiased when using samples of 500 or more and that the asymptotic standard errors are adequate approximations to the variance of the sampling distributions of the parameter estimates. These results are consistent with those reported in Wu and Adams (1993).

In comparing unconditional and conditional estimates of the item parameters, we noted that the use of the collateral information has a negligible effect on the accuracy of the parameter estimates. For both of the simulations that we have reported, we would expect that both the unconditional and the conditional estimates would be consistent. We are yet to undertake a simula-

TABLE 5

A comparison of the regression parameter estimates that result from the two-level conditional model, a two-step approach using plausible values, and a two-step approach using ordinary least squares

Parameter	Generating value	Conditional two-level estimates	Two-step plausible value based estimates	Two-step OLS estimates
β_0	−0.780	−0.782	−0.784	−0.718
β_1	0.657	0.656	0.656	0.508
β_2	−0.002	−0.006	−0.003	−0.005
β_3	0.136	0.135	0.134	0.104
β_4	0.378	0.383	0.385	0.293
R^2	0.326	0.323	0.303	0.247

tion that considers a case where the unconditional model would be expected to produce inconsistent estimates. The conditions under which this will occur are described in Mislevy and Sheehan (1989b).

While it appears that the collateral information plays only a minor role in improving item parameter estimates, it is clear that it can have a significant effect upon EAP ability predictions. Employing the collateral information has the potential to lead to ability predictions that have a substantially smaller mean squared error. Unfortunately, this improvement may be achieved at a cost that many would see as fundamental to equitable measurement—ability predictions that are independent of any influence beyond the individuals responses to the items. In the conditional model, two individuals with identical item response patterns can be assigned different abilities if they do not have identical collateral information.

Finally, we note that while the regression parameter estimates based upon the imputations are almost identical to those produced by the direct estimation, the two-step approaches that use ordinary least squares regression following unconditional ability predictions can produce regression coefficients that are substantially different from those estimated under the two-level model. For simple regression models, traditional attenuation corrections can be applied, but there are no simple corrections for regression coefficients when multiple predictors are used. Additional analyses are necessary to provide a detailed comparison of the direct estimates and those recovered through the use of plausible values. Of particular interest is the choice of the number of imputations to draw. Throughout our work we have followed the NAEP procedures of drawing five plausible values.

Two Step Versus Two Level: A Comparison of Results

We now turn to a comparative analysis of five real data sets collected as part of the Victorian Science Achievement Study that is reported in Adams et al. (1991). In that study, a battery of achievement instruments were administered to random samples of students from Victorian schools. To illustrate the two-level model, we selected five data sets from the study and analyzed the data using three regression methods. The five achievement instruments we used were:

(1) Core, a 15-item multiple-choice test of general science;

(2) Earth and Space, a 9-item partial credit test of student conceptions of the Earth and its place in the solar system;

(3) Force and Motion, a 9-item partial credit test of student conceptions of force and motion;

(4) Matter, a 9-item partial credit test of student conceptions of the structure of matter; and

(5) Light and Sight, a 10-item partial credit test of student conceptions of light and sight.

For each of the five data sets we fitted an unconditional RCML model, produced EAP ability predictions, and then regressed them onto the following four demographic variables:

(1) *SEX*, students' sex, coded 0 for female and 1 for male;

(2) *GRADE*, students' grade level, coded 0 for Grade 5 and 1 for Grade 9;

(3) *SES*, a standardized socioeconomic composite made up of parental occupation and education; and

(4) *GRADE*SEX*, the interaction between *GRADE* and *SEX*.

Second, we fitted the conditional model and obtained direct estimates of the regression parameters. Third, we used a plausible value approach.

The results of these analyses are summarized in Table 6, where for each data set we have reported the regression coefficients and *t*-values that result from fitting each of the three models. The results reported in Table 6 are consistent with those of our simulations. Broadly speaking, the plausible value and two-level conditional approaches produce results that are very similar, while the two-level OLS approach yields regression coefficients for the main effect that appear to be attenuated. For these particular data sets, the attenuation leads to a recognizable underestimation of the direct effects that under some circumstances may lead to important differences in the substantive interpretation of the results. The relationship between the OLS, plausible value, and conditional two-level estimates of the interaction between grade and sex are less clear.

Conclusion

In this article we have shown how structural Rasch models can be viewed as nonlinear multilevel models. This observation invites the possibility of simultaneously modeling the item response process and structural relations

TABLE 6
Regression coefficients and t-values for the science data sets

	Regression coefficients			t-values		
	Two-step OLS	Two-step plausible values	Conditional two-level	Two-step OLS	Two-step plausible values	Conditional two-level
			Core			
Constant	0.135	−0.127	−0.148	0.458	−3.288	−3.863
GRADE	1.069	1.363	1.380	24.009	21.363	23.190
SEX	0.240	0.282	0.299	5.860	4.343	5.631
*GRADE*SEX*	−0.106	−0.088	−0.106	−1.697	−0.864	−1.297
SES	0.349	0.469	0.462	15.811	16.704	15.644
R^2	0.35	0.44	0.46			
			Earth and Space			
Constant	−0.716	−0.746	−0.789	−20.753	−18.173	−16.490
GRADE	0.510	0.631	0.657	9.815	9.783	9.713
SEX	−0.001	−0.052	−0.002	−0.237	−0.816	−0.031
*GRADE*SEX*	0.100	0.193	0.136	1.340	1.858	1.404
SES	0.288	0.376	0.378	10.507	11.731	10.462
R^2	0.24	0.30	0.33			
			Matter			
Constant	−0.999	−1.084	−1.085	−35.080	−29.123	−23.792
GRADE	0.445	0.669	0.650	10.722	12.264	10.614
SEX	−0.023	−0.018	−0.036	−0.601	−0.346	0.619
*GRADE*SEX*	0.030	0.003	0.050	0.529	0.370	0.594
SES	0.142	0.215	0.212	6.949	5.948	6.989
R^2	0.24	0.33	0.39			
			Force and Motion			
Constant	−1.227	−1.307	−1.301	−32.125	−28.811	−21.803
GRADE	0.490	0.654	0.637	8.541	9.188	8.502
SEX	0.017	0.060	0.023	0.313	0.767	0.326
*GRADE*SEX*	0.102	0.091	0.132	1.290	0.294	1.271
SES	0.202	0.258	0.268	7.148	7.230	7.218
R^2	0.19	0.23	0.26			
			Light and Sight			
Constant	−0.446	−0.503	−0.513	−11.069	−9.401	−9.390
GRADE	0.393	0.510	0.517	6.466	6.317	6.447
SEX	0.020	0.033	0.025	0.357	0.419	0.346
*GRADE*SEX*	0.067	0.097	0.096	0.800	0.875	0.876
SES	0.178	0.242	0.236	6.162	5.536	6.171
R^2	0.13	0.17	0.18			

between variables. We have described how the approach, although differently motivated, is formally equivalent to that suggested by Mislevy (1985). We have also applied the method to a different class of measurement models, a class of generalized Rasch models described in Adams and Wilson (1996), which in one sense is more general and in another more restrictive than those that have been presented elsewhere. It is more general in the sense that the model we use is a very general polytomous model that can be applied to a very wide range of measurement contexts. It is more restrictive in the sense that it belongs to the Rasch family.

In examining our approach to estimation, we have also been able to test the suitability of some approximate procedures that are currently employed in NAEP. For example, we have empirically shown that the consecutive estimation of item parameters followed by the population parameters produces item parameter estimates that can probably be treated as equivalent to those produced through a joint analysis. While this is a comforting finding, at least from a practical perspective, additional work does need to be done to formally investigate the relationship between the two different approaches to parameter estimation. Further, in both simulated and real examples, we have shown the similarity of direct estimates of population parameters and those recovered through the use of imputations (plausible values) for the unknown ability values.

The models that we have considered deal only with a unidimensional latent variable. The methods that we employ can, however, be applied to multidimensional extensions of the RCML. The multidimensional extension of the model is reported in Wang (1994) and Adams, Wilson, and Wang (in press), and software for fitting these models with conditioning variables is described by Wu, Adams, and Wilson (1995).

Having taken the multilevel perspective on item response modeling, the obvious next step is to consider the extension of the approach to more than two levels, so that hierarchical sample structures (e.g., students nested within classrooms) can be approxpriately considered. This we plan to do in the near future. Further, we will be considering the possibility of removing the assumptions of normality and considering semiparametric estimation of population distributions.

Note

[1]This clarification is due to an anonymous reviewer.

APPENDIX

In this Appendix we derive the observed information matrix for the model. We use the observed information in Newton steps and in estimating asymptotic standard errors for the maximum likelihood parameter estimates.

To assist in the notation, we introduce the expectations

$$E_{\mathbf{z}}(t|\theta_n) = \Psi(\theta_n, \boldsymbol{\xi}) \sum_{\mathbf{z}\in\boldsymbol{\Omega}} t \exp\{\mathbf{z}'(\mathbf{b}\theta_n + \mathbf{A}\boldsymbol{\xi})\}$$

and

$$E_{\theta_n}(t) = \int_{\theta_n} t\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2 | \mathbf{x}_n)\, d\theta_n,$$

and use $\boldsymbol{\zeta}$ to represent all of the model parameters. $\hat{\boldsymbol{\zeta}}$ are their corresponding maximum likelihood estimates. Similarly, we use $\hat{E}_\mathbf{z}(t/\theta_n)$ and $\hat{E}_{\theta_n}(t)$ to denote the above expectations evaluated at $\hat{\boldsymbol{\zeta}}$.

The following results are useful in deriving the elements of the information matrix:

$$\frac{\partial \log f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi}} = \mathbf{A}'(\mathbf{x}_n - E_\mathbf{z}(\mathbf{z}|\theta_n)), \tag{A1}$$

$$\frac{\partial^2 \log f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi}'\partial \boldsymbol{\xi}} = -\mathbf{A}'(E_\mathbf{z}(\mathbf{z}\mathbf{z}'|\theta_n) - E_\mathbf{z}(\mathbf{z}|\theta_n)E_\mathbf{z}(\mathbf{z}'|\theta_n))\mathbf{A}, \tag{A2}$$

$$\frac{\partial^2 \log f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\beta}'\partial \boldsymbol{\xi}} = 0, \tag{A3}$$

$$\frac{\partial^2 \log f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \sigma^2 \partial \boldsymbol{\xi}} = 0, \tag{A4}$$

$$\begin{aligned}\frac{\partial f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi}'} &= \frac{\partial \log f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi}'} f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)\\ &= (\mathbf{x}_n' - E_\mathbf{z}(\mathbf{z}'|\theta_n))\mathbf{A} f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n) \text{ from (A1)},\end{aligned} \tag{A5}$$

$$\begin{aligned}\frac{\partial f(\mathbf{x}_n; \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\xi}} &= f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)\, \frac{\partial \log f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\xi}'}\\ &= f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2) \int_{\theta_n} \frac{\partial \log f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi}} h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \\ &\qquad \text{from (10)}\\ &= \mathrm{f}(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2) \int_{\theta_n} (\mathbf{x}_n' - E_\mathbf{z}(\mathbf{z}'|\theta_n))\mathbf{A}\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \\ &\qquad \text{from (A1)}\\ &= \mathrm{f}(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)(\mathbf{x}_n' - E_{\theta_n}[E_\mathbf{z}(\mathbf{z}'|\theta_n)])\mathbf{A},\end{aligned} \tag{A6}$$

$$\begin{aligned}\frac{\partial h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)}{\partial \boldsymbol{\xi}'} &= \frac{\partial}{\partial \boldsymbol{\xi}'}\left(\frac{f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n) f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}\right)\\ &= f^{-2}(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)\left\{\frac{\partial f_\mathbf{x}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi}'} f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)\right.\\ &\quad \times f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)\end{aligned}$$

$$- f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n) f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2) \frac{\partial f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\xi}'} \Bigg\}$$

$$= h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n) \Bigg\{ \frac{\partial \log f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi}'}$$

$$- \frac{\partial \log f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\xi}'} \Bigg\}$$

$$= h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)$$

$$\times \Bigg\{ (\mathbf{x}_n' - E_{\mathbf{z}}(\mathbf{z}'|\theta_n))$$

$$- \int_{\theta_n} (\mathbf{x}_n' - E_{\mathbf{z}}(\mathbf{z}'|\theta_n))\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \Bigg\} \mathbf{A}$$

from (A1) and (A6)

$$= h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n) \Bigg\{ \int_{\theta_n} E_{\mathbf{z}}(\mathbf{z}'|\theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$- E_{\mathbf{z}}(\mathbf{z}'|\theta_n) \Bigg\} \mathbf{A}$$

$$= h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n) \{ E_{\theta_n}[E_{\mathbf{z}}(\mathbf{z}'|\theta_n)] - E_{\mathbf{z}}(\mathbf{z}'|\theta_n) \} \mathbf{A}, \tag{A7}$$

$$\frac{\partial \log f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} = \frac{\mathbf{Y}_n}{\sigma^2} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta}), \tag{A8}$$

$$\frac{\partial^2 \log f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}} = -\frac{\mathbf{Y}_n \mathbf{Y}_n'}{\sigma^2}, \tag{A9}$$

$$\frac{\partial^2 \log f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2 \partial \boldsymbol{\beta}} = \frac{\partial}{\partial \sigma^2} \left(\frac{\mathbf{Y}_n}{\sigma^2} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta}) \right)$$

$$= -\frac{\mathbf{Y}_n}{\sigma^4} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta}), \tag{A10}$$

$$\frac{\partial f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}'} = \frac{\partial \log f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}'} f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)$$

$$= f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2) \int_{\theta_n} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta}) \frac{\mathbf{Y}_n'}{\sigma^2} h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

from (11) and (A8)

$$= f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)(E_{\theta_n}(\theta_n) - \mathbf{Y}_n'\boldsymbol{\beta}) \frac{\mathbf{Y}_n'}{\sigma^2}, \tag{A11}$$

$$\frac{\partial h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)}{\partial \boldsymbol{\beta}'} = \frac{\partial}{\partial \boldsymbol{\beta}'}\left(\frac{f_x(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n) f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}\right)$$

$$= f^{-2}(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)\left\{f_x(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n) \frac{\partial f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}'}\right.$$

$$\times f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)$$

$$\left. - f_x(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n) f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2) \frac{\partial f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}'}\right\}$$

$$= h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\left\{\frac{\partial \log f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}'}\right.$$

$$\left. - \frac{\partial \log f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}'}\right\}$$

$$= h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\left\{(\theta_n - \mathbf{Y}_n'\boldsymbol{\beta}) \frac{\mathbf{Y}_n'}{\sigma^2} - (E_{\theta_n}(\theta_n) - \mathbf{Y}_n'\boldsymbol{\beta}) \frac{\mathbf{Y}_n'}{\sigma^2}\right\}$$

from (A8) and (A11)

$$= h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\{\theta_n - E_{\theta_n}(\theta_n)\} \frac{\mathbf{Y}_n'}{\sigma^2}, \tag{A12}$$

$$\frac{\partial \log f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{1}{2\sigma^4}[\sigma^2 - (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2], \tag{A13}$$

$$\frac{\partial^2 \log f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{(\partial \sigma^2)^2} = \frac{1}{2\sigma^6}[\sigma^2 - 2(\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2], \tag{A14}$$

$$\frac{\partial f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = \frac{\partial \log f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)$$

$$= -f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2) \int_{\theta_n} \frac{1}{2\sigma^4}(\sigma^2 - (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2)$$

$$\times h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

from (12) and (A13)

$$= -\frac{f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{2\sigma^4}$$

$$\times \left(\sigma^2 - \int_{\theta_n} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2 h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\right), \tag{A15}$$

and

$$\frac{\partial h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left(\frac{f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n) f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}\right)$$

$$= f^{-2}(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)\left\{f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n) \frac{\partial f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2}\right.$$

$$\times f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)$$

$$\left. - f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n) f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2) \frac{\partial f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2}\right\}$$

$$= h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\left\{\frac{\partial \log f_\theta(\theta_n|\mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2}\right.$$

$$\left. - \frac{\partial \log f(\mathbf{x}_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2}\right\}$$

$$= \frac{h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)}{2\sigma^4}$$

$$\times \left\{(\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2 - \int_{\theta_n} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\right\}. \tag{A16}$$

from (A13) and (A15)

Using the results (A1)–(A16), we are now able to derive the elements of the observed information.

$$\frac{\partial^2 \boldsymbol{\lambda}}{\partial \boldsymbol{\xi}' \partial \boldsymbol{\xi}} = \frac{\partial}{\partial \boldsymbol{\xi}'} \sum_{n=1}^{N} \int_{\theta_n} \frac{\partial \log f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi}} h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$= \sum_{n=1}^{N} \left[\int_{\theta_n} \frac{\partial^2 \log f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'} h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\right]$$

$$+ \sum_{n=1}^{N} \left[\int_{\theta_n} \frac{\partial \log f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi}} \frac{\partial h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)}{\partial \boldsymbol{\xi}'}\, d\theta_n\right]$$

$$= -\mathbf{A}' \sum_{n=1}^{N} \left[\int_{\theta_n} E_{\mathbf{z}}(\mathbf{z}\mathbf{z}'|\theta_n) - E_{\mathbf{z}}(\mathbf{z}|\theta_n)\, E_{\mathbf{z}}(\mathbf{z}'|\theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\right.$$

$$\left. + \int_{\theta_n} (\mathbf{x}_n - E_{\mathbf{z}}(\mathbf{z}|\theta_n))(E_{\mathbf{z}}(\mathbf{z}'|\theta_n) - E_{\theta_n}[E_{\mathbf{z}}(\mathbf{z}'|\theta_n)])\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\right]\mathbf{A}$$

from (A1), (A2) and (A7)

$$= -\mathbf{A}' \sum_{n=1}^{N} \left[\int_{\theta_n} E_{\mathbf{z}}(\mathbf{z}\mathbf{z}'|\theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\right.$$

$$- 2 \int_{\theta_n} E_{\mathbf{z}}(\mathbf{z}|\theta_n)\, E_{\mathbf{z}}(\mathbf{z}'|\theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \tag{A17}$$

$$+ \int_{\theta_n} E_{\mathbf{z}}(\mathbf{z}|\theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \int_{\theta_n} E_{\mathbf{z}}(\mathbf{z}'|\theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \Bigg] \mathbf{A},$$

which, evaluated at the maximum likelihood estimates, gives

$$\left.\frac{\partial^2 \boldsymbol{\lambda}}{\partial \boldsymbol{\xi}' \partial \boldsymbol{\xi}}\right|_{\zeta=\hat{\zeta}} = -\mathbf{A}' \sum_{n=1}^{N} \Bigg[\int_{\theta_n} \hat{E}_{\mathbf{z}}(\mathbf{z}\mathbf{z}'|\theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{x}_n)\, d\theta_n$$

$$- 2 \int_{\theta_n} \hat{E}_{\mathbf{z}}(\mathbf{z}|\theta_n)\, \hat{E}_{\mathbf{z}}(\mathbf{z}'|\theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{x}_n)\, d\theta_n$$

$$+ \hat{E}_{\theta_n}[\hat{E}_{\mathbf{z}}(\mathbf{z}|\theta_n)]\hat{E}_{\theta_n}[\hat{E}_{\mathbf{z}}(\mathbf{z}'|\theta_n)] \Bigg] \mathbf{A}. \tag{A18}$$

$$\frac{\partial^2 \boldsymbol{\lambda}}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}'} \sum_{n=1}^{N} \int_{\theta_n} \frac{\partial \log f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}}\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$= \sum_{n=1}^{N} \Bigg[\int_{\theta_n} \frac{\partial^2 \log f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}}\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$+ \int_{\theta_n} \frac{\partial \log f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} \frac{\partial h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)}{\partial \boldsymbol{\beta}'}\, d\theta_n \Bigg]$$

$$= -\sum_{n=1}^{N} \Bigg[\int_{\theta_n} \frac{\mathbf{Y}_n \mathbf{Y}_n'}{\sigma^2}\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$- \int_{\theta_n} \frac{\mathbf{Y}_n}{\sigma^2} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta}) \{(\theta_n - \mathbf{Y}_n'\boldsymbol{\beta}) - (E_\theta(\theta_n) - \mathbf{Y}_n'\boldsymbol{\beta})\}$$

$$\times h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n) \frac{\mathbf{Y}_n'}{\sigma^2}\, d\theta_n \Bigg], \tag{A19}$$

from (A8), (A9) and (A12)

which, evaluated at the maximum likelihood estimates, gives

$$\left.\frac{\partial^2 \boldsymbol{\lambda}}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}}\right|_{\zeta=\hat{\zeta}} = \sum_{n=1}^{N} \frac{\mathbf{Y}_n}{\hat{\sigma}^2} (\hat{E}_\theta(\theta_n^2) - \hat{E}_\theta(\theta_n)^2) \frac{\mathbf{Y}_n'}{\hat{\sigma}^2} - \frac{\mathbf{Y}_n \mathbf{Y}_n'}{\hat{\sigma}^2}$$

$$= \sum_{n=1}^{N} \frac{\mathbf{Y}_n \mathbf{Y}_n'}{\hat{\sigma}^2} \left[\frac{\hat{E}_\theta(\theta_n^2) - \hat{E}_\theta(\theta_n)^2}{\hat{\sigma}^2} - 1 \right]. \tag{A20}$$

$$\frac{\partial^2 \boldsymbol{\lambda}}{(\partial \sigma^2)^2} = \frac{\partial}{\partial \sigma^2} \sum_{n=1}^{N} \int_{\theta_n} \frac{\partial \log f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2}\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$= \sum_{n=1}^{N} \Bigg[\int_{\theta_n} \frac{\partial^2 \log f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{(\partial \sigma^2)^2}\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$+ \int_{\theta_n} \frac{\partial \log f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\beta}, \sigma^2)}{\partial\sigma^2} \frac{\partial h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)}{\partial\sigma^2} d\theta_n\Bigg]$$

$$= \sum_{n=1}^{N} \Bigg[\int_{\theta_n} \frac{1}{2\sigma^6} (\sigma^2 - 2(\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2)\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$- \frac{1}{4\sigma^8} \int_{\theta_n} (\sigma^2 - (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2)\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)$$

$$\times \left((\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2 - \int_{\theta_n} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\right) d\theta_n\Bigg]$$

from (A13), (A14) and (A16)

$$= \sum_{n=1}^{N} \Bigg[\int_{\theta_n} \frac{1}{2\sigma^6} (\sigma^2 - 2(\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2)\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$+ \frac{1}{4\sigma^8} \int_{\theta_n} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^4\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$- \frac{1}{4\sigma^8} \left(\int_{\theta_n} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\right)^2\Bigg]. \tag{A21}$$

So, using (12) and evaluating at the maximum likelihood estimates, we have

$$\left.\frac{\partial^2 \boldsymbol{\lambda}}{(\partial\sigma^2)^2}\right|_{\zeta=\hat{\zeta}} = -\frac{N}{2\hat{\sigma}^4} + \frac{1}{4\hat{\sigma}^8} \sum_{n=1}^{N} \Bigg[\int_{\theta_n} (\theta_n - \mathbf{Y}_n'\hat{\boldsymbol{\beta}})^4\, h_\theta(\theta_n; \mathbf{Y}_n, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{x}_n)\, d\theta_n$$

$$-\left(\int_{\theta_n} (\theta_n - \mathbf{Y}_n'\hat{\boldsymbol{\beta}})^2\, h_\theta(\theta_n; \mathbf{Y}_n, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{x}_n)\, d\theta_n\right)^2\Bigg]. \tag{A22}$$

$$\frac{\partial^2 \boldsymbol{\lambda}}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\xi}} = \frac{\partial}{\partial\boldsymbol{\beta}'} \sum_{n=1}^{N} \int_{\theta_n} \frac{\partial \log f_x(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial\boldsymbol{\xi}} h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$= \sum_{n=1}^{N} \Bigg[\int_{\theta_n} \frac{\partial \log f_x(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\xi}} h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$+ \int_{\theta_n} \frac{\partial \log f_x(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial\boldsymbol{\xi}} \frac{h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)}{\partial\boldsymbol{\beta}'} d\theta_n\Bigg]$$

$$= \mathbf{A}' \sum_{n=1}^{N} \Bigg[\int_{\theta_n} (\mathbf{x}_n - E_z(\mathbf{z}|\theta_n))\, (\theta_n - E_\theta(\theta_n))\, \frac{\mathbf{Y}_n'}{\sigma^2} h_\theta(\theta_n; \mathbf{Y}_n; \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n\Bigg], \tag{A23}$$

from (A1), (A3) and (A12)

which, evaluated at the maximum likelihood estimates, gives

$$\left.\frac{\partial^2 \boldsymbol{\lambda}}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\xi}}\right|_{\zeta=\hat{\zeta}} = -\mathbf{A}' \sum_{n=1}^{N} \left(\int_{\theta_n} \theta_n \hat{E}_{\mathbf{z}}(\mathbf{z}|\theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{x}_n)\, d\theta_n - \hat{E}_\theta[\hat{E}_{\mathbf{z}}(\mathbf{z}|\theta_n)\hat{E}_\theta(\theta_n)] \right) \frac{\mathbf{Y}_n'}{\hat{\sigma}^2}. \tag{A24}$$

$$\frac{\partial^2 \boldsymbol{\lambda}}{\partial \sigma^2 \partial \boldsymbol{\xi}} = \frac{\partial}{\partial \sigma^2} \sum_{n=1}^{N} \int_{\theta_n} \frac{\partial \log f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi}} h(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$= \sum_{n=1}^{N} \left[\int_{\theta_n} \frac{\partial^2 \log f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \sigma^2\, \partial \boldsymbol{\xi}} h(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \right.$$

$$\left. + \int_{\theta_n} \frac{\partial \log f_{\mathbf{x}}(\mathbf{x}_n; \boldsymbol{\xi}|\theta_n)}{\partial \boldsymbol{\xi}} \frac{\partial h(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)}{\partial \sigma^2}\, d\theta_n \right]$$

from (A1), (A4) and (A16)

$$= \frac{\mathbf{A}'}{2\sigma^4} \sum_{n=1}^{N} \left[\int_{\theta_n} (\mathbf{x}_n - E_{\mathbf{z}}(\mathbf{z}|\theta_n)) \left((\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2 \right. \right.$$

$$\left. - \int_{\theta_n} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \right)$$

$$\left. \times\, h(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \right], \tag{A25}$$

which, evaluated at the maximum likelihood estimates, gives

$$\left.\frac{\partial^2 \boldsymbol{\lambda}}{\partial \sigma^2 \partial \boldsymbol{\xi}}\right|_{\zeta=\hat{\zeta}} = -\frac{\mathbf{A}'}{2\sigma^4} \sum_{n=1}^{N} \left[\int_{\theta_n} \hat{E}_{\mathbf{z}}(\mathbf{z}|\theta_n)(\theta_n - \mathbf{Y}_n'\hat{\boldsymbol{\beta}})^2\, h_\theta(\theta_n; \mathbf{Y}_n, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{x}_n)\, d\theta_n \right.$$

$$- \int_{\theta_n} \hat{E}_{\mathbf{z}}(\mathbf{z}|\theta_n)\, h_\theta(\theta_n; \mathbf{Y}_n, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{x}_n)\, d\theta_n$$

$$\left. \times \int_{\theta_n} (\theta_n - \mathbf{Y}_n'\hat{\boldsymbol{\beta}})^2\, h_\theta(\theta_n; \mathbf{Y}_n, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{x}_n)\, d\theta_n \right]. \tag{A26}$$

$$\frac{\partial^2 \boldsymbol{\lambda}}{\partial \sigma^2 \partial \boldsymbol{\beta}} = \frac{\partial}{\partial \sigma^2} \sum_{n=1}^{N} \int_{\theta_n} \frac{\partial \log f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} h(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n$$

$$= \sum_{n=1}^{N} \left[\int_{\theta_n} \frac{\partial^2 \log f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2 \partial \boldsymbol{\beta}} h(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \right.$$

$$\left. + \int_{\theta_n} \frac{\partial \log f_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} \frac{\partial h(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)}{\partial \sigma^2}\, d\theta_n \right]$$

from (A8), (A10) and (A16)

$$= \sum_{n=1}^{N} \left[\int_{\theta_n} \frac{\mathbf{Y}_n}{\sigma^4} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})\, h(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \right.$$

$$+ \int_{\theta_n} \left\{ \frac{\mathbf{Y}_n}{2\sigma^6} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta}) \left((\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2 - \int_{\theta_n} (\theta_n - \mathbf{Y}_n'\boldsymbol{\beta})^2\, h_\theta(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n)\, d\theta_n \right) \right.$$

$$\left. \left. \times\, h(\theta_n; \mathbf{Y}_n, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2|\mathbf{x}_n) \right\} d\theta_n \right]. \tag{A27}$$

So, using (11) and evaluating at the maximum likelihood estimates, we get

$$\left. \frac{\partial^2 \boldsymbol{\lambda}}{\partial \sigma^2 \partial \boldsymbol{\beta}} \right|_{\zeta=\hat{\zeta}} = -\frac{1}{2\hat{\sigma}^6} \sum_{n=1}^{N} \left[\int_{\theta_n} \mathbf{Y}_n(\theta_n - \mathbf{Y}_n'\hat{\boldsymbol{\beta}})^3\, h_\theta(\theta_n; \mathbf{Y}_n, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{x}_n)\, d\theta_n \right.$$

$$- \int_{\theta_n} \mathbf{Y}_n(\theta_n - \mathbf{Y}_n'\hat{\boldsymbol{\beta}})\, h_\theta(\theta_n; \mathbf{Y}_n, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{x}_n)\, d\theta_n$$

$$\left. \times \int_{\theta_n} (\theta_n - \mathbf{Y}_n'\hat{\boldsymbol{\beta}})^2\, h_\theta(\theta_n; \mathbf{Y}_n, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{x}_n)\, d\theta_n \right]. \tag{A28}$$

References

Adams, R. J. (1989). *Estimating measurement error and its effects on statistical analysis.* Unpublished doctoral dissertation, University of Chicago.

Adams, R. J., Doig, B. A., & Rosier, M. (1991). *Science learning in Victorian schools* (Research Monograph 41). Hawthorn, Australia: Australian Council for Educational Research.

Adams, R. J., & Wilson, M. R. (1996). A random coefficients multinomial logit: A generalized approach to fitting Rasch models. In G. Engelhard & M. Wilson (Eds.), *Objective measurement III: Theory into practice* (pp. 143–166). Norwood, NJ: Ablex.

Adams, R. J., Wilson, M. R., & Wang, W. C. (in press). The multidimensional random coefficients multinomial logit. *Applied Psychological Measurement.*

Anderson, D. A., & Aitkin, M. (1985). Variance component models with binary response: Interviewer variability. *Journal of the Royal Statistical Society, Series B, 47,* 203–210.

Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561–573.

Beaton, A. E. (1987). *Implementing the new design: The NAEP 1983–84 technical report* (Rep. No. 15-TR-20). Princeton, NJ: Educational Testing Service.

Bock, R. D. (1989). *Multilevel analysis of educational data.* London: Academic Press.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika, 46,* 443–459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35,* 179–187.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101,* 147–158.

Cressie, N., & Holland, P. W. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika, 48*, 129–141.

de Leeuw, J., & Verhelst, N. (1986). Maximum likelihood estimation in generalized Rasch models. *Journal of Educational Statistics, 11*, 183–196.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1–38.

Fischer, G. H. (1983). Logistic latent trait models with linear constraints. *Psychometrika, 48*, 3–26.

Goldstein, H. (1987). *Multilevel models in education and social research.* London: Charles Griffin and Co.

Goldstein, H. (1991). Non-linear multi-level models, with an application to discrete response data. *Biometrika, 78*, 45–51.

Holland, P. W. (1990). On the sampling theory foundation of item response theory models. *Psychometrika, 55*, 577–601.

Johnson, E. G., Mazzeo, J., & Kline, D. (1993). *Technical report of the NAEP 1992 trial state assessment program in mathematics* (Rep. No. 23-ST05). Princeton, NJ: Educational Testing Service.

Kelderman, H. (1989, March). *Loglinear multidimensional IRT models for polytomously scored items.* Paper presented at the Fifth International Objective Measurement Workshop, University of California, Berkeley.

Linacre, J. M. (1989). *Many faceted Rasch measurement.* Unpublished doctoral dissertation, University of Chicago.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*, 359–381.

Mislevy, R. J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association, 80*, 993–997.

Mislevy, R. J. (1987). Exploiting auxiliary information about examinees in the estimation of item parameters. *Applied Psychological Measurement, 11*, 81–91.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*, 177–196.

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29*, 133–161.

Mislevy, R. J., & Bock, R. D. (1983). *BILOG: Analysis and scoring of binary items with one-, two-, and three-parameter logistic models.* Mooresville, IN: Scientific Software Inc.

Mislevy, R. J., & Sheehan, K. M. (1989a). Information matrices in latent variable models. *Journal of Educational Statistics, 14*, 335–350.

Mislevy, R. J., & Sheehan, K. M. (1989b). The role of collateral information about examinees in item parameter estimation. *Psychometrika, 54*, 661–679.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159–176.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). Chicago: University of Chicago Press.

Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics, 13*, 85–116.

Raudenbush, S. W., & Bryk, A. S. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics, 10*, 75–98.

Stiratelli, R., Laird, N., & Waire, J. H. (1984). Random effects models for serial observations with binary response. *Biometrics, 40*, 961–971.

Thomas, N. (1992). Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *The Journal of Computational and Graphical Statistics, 2*, 309–322.

Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scandinavian Journal of Statistics, 9*, 23–30.

Wang, W. C. (1994). *Implementation and application of the multidimensional random coefficients multinomial logit.* Unpublished doctoral dissertation, University of California, Berkeley.

Wilson, M. R. (1992). The ordered partition model: An extension of the partial credit model. *Applied Psychological Measurement, 16*, 309–325.

Wilson, M. R., & Adams, R. J. (1993). Marginal maximum likelihood estimation for the ordered partition model. *Journal of Educational Statistics, 18*, 69–90.

Wilson, M. R., & Adams, R. J. (1995). Rasch models for item bundles. *Psychometrika, 60*, 181–198.

Wong, G. Y., & Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association, 80*, 513–524.

Wu, M. L., & Adams, R. J. (1993, April). *Simulating parameter recovery for the random coefficients multinomial logit.* Paper presented at the Seventh International Objective Measurement Workshop, Atlanta, GA.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1995, April). *MATS: Multi-aspect test software.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Zwick, R. (Ed.). (1992). National Assessment of Educational Progress [Special issue]. *Journal of Educational Statistics, 17*(2).

Authors

RAYMOND J. ADAMS is Principal Research Fellow, Australian Council for Educational Research, Private Bag 55, Camberwell, Victoria 3124, Australia; adams@acer.edu.au. He specializes in psychometrics and multilevel modeling.

MARK WILSON is Professor, Graduate School of Education, Tolman Hall, University of California, Berkeley, CA 94720; mrwilson@violet.berkeley.edu. He specializes in psychometrics and educational statistics.

MARGARET WU is Research Fellow, Australian Council for Educational Research, Private Bag 55, Camberwell, Victoria 3124, Australia; wu@acer.edu.au. She specializes in psychometrics and multilevel modeling.

Received November 9, 1993
Revision received January 11, 1995
Second revision received June 26, 1995
Third revision received August 8, 1995
Accepted September 20, 1995