RR-88-54-ONR

4

RANDOMIZATION-BASED INFERENCES ABOUT LATENT VARIABLES FROM COMPLEX SAMPLES

Robert J. Mislevy

This research was sponsored in part by the
Office for Educational Research and Improvement
Center for Education Statistics, under
Grant No. NIE-G-83-0011; and the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-88-K-0304
R&T 4421552

Robert J. Mislevy, Principal Investigator

Educational Testing Service
Princeton, New Jersey

September 1988

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION
Unclassified

1b RESTRICTIVE MARKINGS

2a SECURITY CLASSIFICATION AUTHORITY

2b DECLASSIFICATION / DOWNGRADING SCHEDULE

3 DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release;
distribution unlimited.

4 PERFORMING ORGANIZATION REPORT NUMBER(S)
RR-88-54-ONR

5 MONITORING ORGANIZATION REPORT NUMBER(S)

6a NAME OF PERFORMING ORGANIZATION
Educational Testing Service

6b OFFICE SYMBOL (If applicable)

7a NAME OF MONITORING ORGANIZATION
Cognitive Science Program, Office of Naval Research (Code 1142CS), 800 North Quincy Street

6c ADDRESS (City, State, and ZIP Code)
Princeton, NJ 08541

7b ADDRESS (City, State, and ZIP Code)
Arlington, VA 22217-5000

8a NAME OF FUNDING / SPONSORING ORGANIZATION

8b OFFICE SYMBOL (If applicable)

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
N00014-88-K-0304

8c ADDRESS (City, State, and ZIP Code)

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61153N	RR04204	RR04204-01	R&T4421552

11 TITLE (Include Security Classification)
Randomization-Based Inferences about Latent Variables from Complex Samples (Unclassified)

12 PERSONAL AUTHOR(S)
Robert J. Mislevy

13a TYPE OF REPORT
Technical

13b TIME COVERED
FROM ________ TO ______

14 DATE OF REPORT (Year, Month, Day)
September 1988

15 PAGE COUNT
64

16 SUPPLEMENTARY NOTATION

17 COSATI CODES

FIELD	GROUP	SUB-GROUP
05	10	

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Complex samples, item response theory, latent structure, missing data, multiple imputation, National Assessment of Educational Progress, sample surveys

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

the

Standard procedures for drawing inferences from complex samples do not apply when the variable of interest θ cannot be observed directly, but must be inferred from the values of secondary random variables that depend on θ stochastically. Examples are examinee proficiency variables in item response theory models and class memberships in latent class models. This paper uses Rubin's "multiple imputation" approach to approximate sample statistics that would have been obtained, had θ been observable. Associated variance estimates account for uncertainty due to both the sampling of respondents from the population and the latency of θ. The approach is illustrated with artificial examples and with data from the 1984 National Assessment for Educational Progress reading survey. (KS)

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL
Dr. Charles E. Davis

22b TELEPHONE (Include Area Code)
202-696-4046

22c OFFICE SYMBOL
ONR 1142CS

Previous editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

Unclassified

Randomization-Based Inferences about Latent Variables from Complex Samples

Robert J. Mislevy

Educational Testing Service

September 1988

This research was supported by Grant No. NIE-G-83-0011 of the Office for Educational Research and Improvement, Center for Education Statistics, and Contract No. N00014-88-K-0304, R&T 4421552 from the Cognitive Sciences Program, Cognitive and Neural Sciences Division, Office of Naval Research. It does not necessarily reflect the views of either agency. I am grateful to R. Darrell Bock for calling my attention to the applicability of Rubin's multiple imputations to the assessment setting; to Albert Beaton and Eugene Johnson for enlightening discussions on the topic; and to Henry Braun, Ben King, Debra Kline, Paul Rosenbaum, Don Rubin, John Tukey, Kentaro Yamamoto, and Rebecca Zwick for comments on earlier drafts. Example 3 is based on the analysis of the 1984 National Assessment for Educational Progress reading survey, carried out at Educational Testing Service through the tireless efforts of too many people to mention by name, under the direction of Albert Beaton, Director of NAEP Data Analyses.

Randomization-based Inferences about Latent Variables from Complex Samples

Abstract

Standard procedures for drawing inferences from complex samples do not apply when the variable of interest θ cannot be observed directly, but must be inferred from the values of secondary random variables that depend on θ stochastically. Examples are examinee proficiency variables in item response theory models and class memberships in latent class models. This paper uses Rubin's "multiple imputation" approach to approximate sample statistics that would have been obtained, had θ been observable. Associated variance estimates account for uncertainty due to both the sampling of respondents from the population and the latency of θ. The approach is illustrated with artificial examples and with data from the 1984 National Assessment for Educational Progress reading survey.

Key words: Complex samples, item response theory, latent structure, missing data, multiple imputation, National Assessment of Educational Progress, sample surveys.

Introduction

Latent-variable models are used in the social sciences to provide parsimonious explanations of associations among observed variables in terms of theoretical constructs. Practical benefits can accrue as well, as when examinees who have been presented different test items are compared using item response theory (IRT) psychometric models, or when consumer satisfaction levels are tracked over time with different survey questions by means of a latent class model.

This paper addresses the problem of estimating the distributions of latent variables in finite populations, when the data are obtained in complex sampling designs. The solution it offers is to apply Rubin's (1987) "multiple imputations" procedures for missing data in sample surveys. This approach provides consistent estimates of population characteristics, supports statements of precision that account for both the sampling of subjects and the latency of variables, and produces filled-in pseudo datasets that are easy for secondary researchers to analyze correctly.

The following sections briefly review randomization-based inference about manifest variables from complex samples, and multiple imputation for nonresponse. The next sections apply these ideas to latent-variable measurement models and illustrate them with an example using classical test theory. Computing approximations are then discussed. The paper concludes by sketching the implementation, the results, and the lessons learned in applying the procedures to the 1984 reading survey of the National Assessment of Educational Progress (NAEP).

Drawing Inferences from Complex Samples

Standard analyses of sample survey data (e.g., Cochran, 1977) provide estimates of population characteristics and associated sampling variances when the values of survey variables are observed from each respondent in the realized sample. This section gives background and notation for the analysis of survey data, first when all responses are in fact observed, then when some are missing.

Randomization-Based Inference with Complete Data

Consider a population of N identifiable units, indexed by the subscript i. Associated with each unit are two possibly vector-valued variables Y and Z. The values of the design variables, Z, are known for all units before observations are made, but the values of the survey variables, Y, are not. Let $(\underset{\sim}{X},\underset{\sim}{Z})$ denote the population matrix of values. Interest lies in the value of a function $S = S(\underset{\sim}{Y},\underset{\sim}{Z})$ of the population values; examples include a population total for an element of Y, a subpopulation mean for an element of Y given specified values of elements of Z, and the linear regression coefficients of some elements of Y on others.

Any of these functions could be calculated directly if values of Y and Z were observed for all units in the population.

There are 2^N possible subsets of units, and a sample design assigns a probability to each. A simple random sample of size 100, for example, assigns equal probability to all subsets of 100 units, and zero to all other subsets. A sample design yields a "complex sample" when it exhibits one or more of the following features: unequal probabilities of selection for different units; stratification, which ensures prespecified rates of representation in the sample according to values of Z; or clustering, which uses values of Z to link selection probabilities of units when their joint occurrence facilitates gathering the data (e.g., it is easier to interview two respondents in the same town than two in different towns).

Randomization-based inference about S is based on the distribution of a statistic $s = s(\underset{\sim}{y})$ calculated using $\underset{\sim}{y}$, the Y values of a subset of units selected in accordance with a prespecified sample design. In practical work, sample designs are usually constructed so as to yield a nearly unbiased statistic s and an estimate of its variance in the form of another statistic $U = U(\underset{\sim}{y})$. Inferences are then drawn using the normal approximation

$$(s - S) \sim N(0,U) .$$

Under the randomization approach to inference in sample surveys, population values $\underset{\sim}{Y}$ are taken as fixed unknown quantities, and the notion of randomness enters only in the selection of a sample in accordance with the sample design. Under the model-based approach, in contrast, $\underset{\sim}{Y}$ is viewed as a realized sample from a hypothetical "superpopulation" under which variables are distributed in accordance with some model $p(y|z)$ (Cassel, Sarndal, and Wretman, 1977). The randomization approach dominates current practice, and will be used in the sequel to deal with uncertainty due to sampling subjects. It will be seen that superpopulation concepts are required nonetheless to handle missingness and latency.

Randomization-Based Inference with Incomplete Data

In practice, values of one or more survey variables will not be observed from some subjects in the realized sample, for reasons that may or may not be related to the values that would have been observed. For any respondent, let the partitioning $y=(y_{mis},y_{obs})$ indicate the elements of the survey variables that were missing and observed. Much progress has been made extending inferential procedures to survey data with missing responses when those responses are <u>missing at random</u> (MAR); that is, the probability that the elements in y_{mis} are missing may depend on the values of y_{obs} and z, but beyond that, not on the value of y_{mis} (Little and Rubin, 1987; Rubin, 1976).

Let $(\underset{\sim}{y}_{mis},\underset{\sim}{y}_{obs})$ be the matrices of missing and observed survey variables in a realized sample. The values of $\underset{\sim}{y}_{obs}$--which may comprise different elements of y for different subjects--are now known, but the values of $\underset{\sim}{y}_{mis}$ are not. It is not possible to calculate s directly, but it may be possible to calculate its conditional expectation:

$$E[s(\underset{\sim}{y})|\underset{\sim}{y}_{obs}] = \int s(\underset{\sim}{y}_{mis},\underset{\sim}{y}_{obs})\ p(\underset{\sim}{y}_{mis}|\underset{\sim}{y}_{obs},\underset{\sim}{z})\ d\underset{\sim}{y}_{mis}\ . \quad (1)$$

The predictive density $p(\underset{\sim}{y}_{mis}|\underset{\sim}{y}_{obs},\underset{\sim}{z})$ expresses the extent of knowledge about what the missing responses might have been, given the observed responses and the survey variables. If MAR holds, the predictive distribution of a missing element for subject i--say $y_{i,mis}$--is approximated by the responses to that element from subjects who did respond to it <u>and</u> have the same z values as subject i and the same responses on the elements in $y_{i,obs}$. In large surveys with few missing values, one can use these empirical distributions directly. The Census Bureau's "hot deck" imputation procedure (see Ford, 1983), for example, assumes that predictive densities are independent over respondents--i.e.,

$$p(\underset{\sim}{y}_{mis}|\underset{\sim}{y}_{obs},\underset{\sim}{z}) = \prod_i p(y_{i,mis}|y_{i,obs},z_i)\ \text{--}$$

and calculates s with each person's observed responses and, for those he or she is missing, draws from the actual responses of suitably similar respondents.

Alternatively, one can posit a functional form for $p(y_{mis}|y_{obs},z)$, such as a regression model with parameters β. It is usually reasonable to assume independence over respondents, but now conditional on the unknown parameters β; that is,

$$p(\underset{\sim}{y}_{mis}|\underset{\sim}{y}_{obs},\underset{\sim}{z},\beta) = \prod_{i} p(y_{i,mis}|y_{i,obs},z_{i},\beta) \quad . \qquad (2)$$

In this case, the appropriate predictive distribution is obtained after marginalizing with respect to β:

$$p(\underset{\sim}{y}_{mis}|\underset{\sim}{y}_{obs},\underset{\sim}{z}) = \int p(\underset{\sim}{y}_{mis}|\underset{\sim}{y}_{obs},\underset{\sim}{z},\beta)\; p(\beta|\underset{\sim}{y}_{obs})\; d\beta \quad , \qquad (3)$$

where $p(\beta|\underset{\sim}{y}_{obs})$ is the posterior distribution of β after $\underset{\sim}{y}_{obs}$ has been observed.

Multiple Imputation for Missing Responses

As with the hot deck, one can obtain a rough numerical approximation of (1)--an unbiased estimate of the expectation of s--by filling in each missing response with a random draw from its predictive density, then calculating s as if these imputations were the true response values. An analogous approximation of U using these single imputations underestimates the variability of the resulting estimate of S, however. It accounts for uncertainty due to sampling subjects, but not uncertainty due to imputing values for missing responses. To remedy this deficiency, Rubin suggests <u>multiple</u> imputations. When the statistic s is scalar and the predictive distribution for $\underset{\sim}{y}_{mis}$ is model-based, one proceeds as follows:

1. Determine the posterior distribution $p(\beta|\underset{\sim}{y}_{obs})$.

2. Produce M distinct "completed" datasets $\underset{\sim}{y}_{(m)}$, m=1,...,M. Each looks like a complete dataset; it has the values $\underset{\sim}{y}_{obs}$ for the responses that were observed, and, for those that were not, a draw from (3). Two steps complete $\underset{\sim}{y}_{(m)}$:

   a. Draw a value $\beta_{(m)}$ from $p(\beta|\underset{\sim}{y}_{obs})$.

   b. For each respondent with one or more missing responses, draw a value from $p(y_{i,mis}|y_{i,obs},z_i,\beta=\beta_{(m)})$. Taken togethe⌄, the resulting imputation for $y_{i,mis}$ and the observed value of $y_{i,obs}$ constitute $y_{i(m)}$, the "completed" response of subject i.

3. Using each completed dataset, calculate $s_{(m)}=s(\underset{\sim}{y}_{(m)})$ and $U_{(m)}=U(\underset{\sim}{y}_{(m)})$.

4. The final estimate of S is the average of the M estimates from the completed datasets:

$$s_M = \sum_{m=1}^{M} s_{(m)} \;/\; M \quad . \qquad (4)$$

5. The estimated sampling variance of s_M as an estimate of S, say V_M, is the sum of two components:

$$V_M = U_M + (1+M^{-1})\ B_M\ , \qquad (5)$$

where

$$U_M = \sum_{m=1}^{M} U_{(m)}\ /\ M$$

quantifies uncertainty due to sampling subjects, and

$$B_M = \sum_{m=1}^{M} (s_{(m)} - s_M)^2\ /\ (M-1)\ ,$$

the variance among the estimates of s from the M completed datasets, quantifies uncertainty due to missing responses from the sampled subjects.

6. If inferences about S would have been based on $(s-S) \sim N(0,U)$ had all responses been observed, inferences are now based on

$$(s_M-S) \sim t_\nu(0,V_M)\ , \qquad (6)$$

a t-distribution, with degrees of freedom given by

$$\nu = (M-1)\ (1+r_M^{-1})^2\ ,$$

where r_M is the proportional increase in variance due to missingness:

$$r_M = (1+M^{-1})\ B_M\ /\ U_M\ .$$

When B_M is small relative to U_M, ν is large and the normal approximation to the t-distribution suffices.

For k-dimensional s, such as the vector of coefficients in a multiple regression analysis, each $U_{(m)}$ and U_M are covariance matrices, and B_M is an average of squares and cross products rather than simply an average of squares. Then the quantity

$$(S-s_M)\ V^{-1}\ (S-s_M)' \tag{7}$$

is approximately F-distributed with k and ν degrees of freedom, with ν defined as above but with a matrix generalization of r_M:

$$r_M = (1+M^{-1})\ \mathrm{Trace}(B_M U_M^{-1})/k\ .$$

By the same reasoning as used for the normal approximation for scalar s, a chi-square distribution on k degrees of freedom often suffices.

Example 1: A Numerical Illustration

Consider a large population in which the scalar y is distributed $N(\mu,1)$, and it is desired to estimate μ from an SRS of size 12. If y values for all 12 units in the realized sample were observed, inferences about μ would be based on $\bar{y}$, the sample mean, and U, the squared standard error of the mean (SEM^2). U is 1/N since the population variance is known to be one. In particular,

$(\bar{y}-\mu) \sim N(0,1/12)$. Suppose, though, that the values for the last two sampled units are missing (at random). The first ten are observed to take the values shown below:

-.797, -.176, 1.419, .029, -1.107,
1.794, -1.619, 1.104, .418, .161.

These observations comprise $\underset{\sim}{y}_{obs}$, while $\underset{\sim}{y}_{mis}$ is (y_{11},y_{12}). Using multiple imputations to estimate the population mean is accomplished as follows.

Since we are assuming MAR and there are no collateral or design variables, the distributions of y_{11} and y_{12} are simply $N(\mu,1)$ too. What we know about μ from the first ten observations is conveyed by their mean and the SEM^2; with an indifference prior, $p(\mu|\underset{\sim}{y}_{obs})$ is $N(.123,.100)$.

A completed dataset consists of the ten observed y values and an imputation for (y_{11},y_{12}). For the sake of illustration, five completed datasets will be constructed. The procedure for each consists of two steps:

a. A value $\mu_{(m)}$ is drawn from $p(\mu|\underset{\sim}{y}_{obs})$, or $N(.123,.1)$.

b. Imputations $y_{11(m)}$ and $y_{12(m)}$ are drawn from $N(\mu_{(m)},1)$.

A table of random normal deviates gave the following results:

m	$\mu_{(m)}$	$y_{11(m)}$	$y_{12(m)}$
1	.306	-.095	-.801
2	.238	-.599	.644
3	.033	-.029	.510
4	.089	.386	-1.248
5	.001	.205	2.106

From each completed dataset, an estimate $\bar{y}_{(m)}$ is calculated, the mean of ten observed y values and two imputations. The results are .028, .106, .143, .031, and .295. For each m, $U_{(m)}$=1/12, an SEM^2 appropriate for twelve observations from a normal population with a variance known to be one.

The estimate of μ based on the imputed values $\bar{y}_M$ is the average of the five completed-data estimates, or .121. Not surprisingly, this is close to the estimate .123 based on $\underset{\sim}{y}_{obs}$, since that is its expected value. Indeed, y_M would converge to .123 as M increased. The estimated sampling variance of $\bar{y}_M$ is obtained by (5) as

$$
\begin{aligned}
V_M &= \Sigma\, U_{(m)}/5 + (1+5^{-1})\ \Sigma\ (\bar{y}_{(m)}-\bar{y}_M)^2/4 \\
&= 1/12 + 1.2 \times .012 \\
&= .098.
\end{aligned}
$$

This value is very close to 1/10, the SEM^2 that corresponds to a sample of size 10--as it should, since this is what the data actually are.

Randomization-based inferences about μ using the multiple imputations are based on

$$(.121 - \mu) \sim t_{184}(0,.098) \ ,$$

a t-distribution with degrees of freedom obtained by (6) with r_M, the proportional increase in variance due to the missingness of y_{11} and y_{12}, equal to .173.

Latent Variables and Sample Surveys

This section provides notation and background for latent-variable models, and shows how they can be handled in complex samples with multiple imputations.

Latent-Variable Models

Latent variables in educational and psychological measurement models account for regularities in observable data; for example, examinees' tendencies to give correct responses to test items, or respondents' inclinations toward liberal responses on social questions. The probability that a subject with latent parameter θ will make the response $x_{(j)}$ to Item j is modeled as a function of θ and a possibly vector-valued item parameter β_j, as $p(x_{(j)}|\theta,\beta_j)$. The assumption of local or conditional independence posits that the latent variable accounts for associations among responses to various items in a specified domain; i.e., if $x = (x_{(1)},...,x_{(n)})$ is a vector of responses to n items, then

$$p(x|\theta,\beta) = \prod_{j}^{n} p(x_{(j)}|\theta,\beta_j) \ , \tag{8}$$

where $\beta = (\beta_1,...,\beta_n)$. Moreover, the latent variable is typically assumed to account for associations between response variables and collateral subject variables such as demographic or educational standing. Denoting collateral variables by y and z to anticipate developments below, the extension of local independence posits

$$p(x|\theta,\beta,y,z) = p(x|\theta,\beta) \ . \tag{9}$$

When such a model is found to provide an adequate fit to data, observing the responses to any subset of items induces a likelihood function for θ through (8), and it becomes possible to draw inferences about individual values of θ or about their distributions in populations even though different subjects respond to different items. This capability is particularly attractive in educational measurement, as when an IRT model can be used to customize tests to examinees adaptively, or to update the item pools in educational assessments over time.

If the focus is on measuring individuals for placement or selection decisions, enough items can be administered to each to make the likelihood function for his or her θ peak sharply. A satisfactorily precise point estimate of each θ, such as the MLE $\hat{\theta}$ or the Bayes mean estimate $\tilde{\theta} = E(\theta|x)$, can then be obtained. If the focus is on the parameters α of a population distribution of θ, however, these locally optimal point estimates can be decidedly nonoptimal. For a fixed test of fixed length, their distributions do not converge to the true distribution of θ as examinee samples increase. It becomes necessary to estimate α directly, bypassing the intermediary stage of calculating point estimates for individuals.

Suppose the data matrix $\underset{\sim}{x}$ consisted of response vectors $(x_1, \ldots, x_N)$ of an SRS of respondents. The starting point for the direct estimation of the parameters α of the density function $p(\theta|\alpha)$ in this context would be the marginal probability:

$$p(\underset{\sim}{x}|\alpha,\beta) = \prod_{i=1}^{N} \int p(x_i|\theta,\beta)\, p(\theta|\alpha)\, d\theta \; . \qquad (10)$$

A number of recent papers in the statistical and psychometric literature show how to obtain Bayesian or maximum likelihood estimates of α and/or β using (10) (e.g., Andersen and Madsen, 1977; Bock and Aitkin, 1981; Dempster, Laird, and Rubin, 1977; Laird, 1978; Mislevy, 1984, 1985; Sanathanan and Blumenthal, 1978; and Rigdon and Tsutakawa, 1983). As they are presented, however, these methods are poorly suited to general analyses of survey data involving latent variables. As well as being limited to SRS data, they require advanced statistical concepts and iterative computing methodologies that render them inaccessible to the typical secondary analyst.

Latent Variables as Missing Responses

A key insight for dealing with latent variables in sample surveys is to view them as survey variables whose responses are missing for every respondent (e.g., Dempster, Laird, and Rubin, 1977, on factor analysis models). Their missingness satisfies MAR because they are missing regardless of their values, and as such are amenable to the (relatively) simple procedures for MAR missing data described above. In essence, knowledge about subjects' latent variables θ will be represented in the form of predictive distributions conditional on what can be observed--background survey variables y, design variables z, and item responses x whose distributions are assumed to be governed by θ.

Suppose the object of inference is the scalar $S = S(\underset{\sim}{\Theta},\underset{\sim}{X},\underset{\sim}{Y},\underset{\sim}{Z})$, some function of the population values of latent variables, item response variables, background survey variables, or design variables of all units. Suppose further that a sample design has been specified. Three assumptions are central to drawing randomization-based inferences about S. The first is that one would know how to proceed if values of θ were observed rather than latent:

Assumption 1. If values of θ could be observed from sampled respondents, along with values of x and y, a sample statistic $s = s(\underset{\sim}{\theta},\underset{\sim}{x},\underset{\sim}{y})$ and an associated variance estimator $U = U(\underset{\sim}{\theta},\underset{\sim}{x},\underset{\sim}{y})$ would be available for randomization-based inference about S, via $(s-S) \sim N(0,U)$.

Inferences about S cannot be based on a direct calculation of s because all values of θ are missing. As before, however, one can base inferences on its conditional expectation given what is known--the sample values $\underset{\sim}{x}$ and $\underset{\sim}{y}$ and the population values $\underset{\sim}{Z}$--by adapting (1) as follows:

$$E(s|\underset{\sim}{x},\underset{\sim}{y}) = \int s(\underset{\sim}{\theta},\underset{\sim}{x},\underset{\sim}{y})\ p(\underset{\sim}{\theta}|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{Z})\ d\underset{\sim}{\theta}\ . \qquad (11)$$

The second and third assumptions are needed to define the predictive distribution for $\underset{\sim}{\theta}$ that appears in (11), namely $p(\underset{\sim}{\theta}|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{Z})$. These assumptions are embodied in the latent variable model and a superpopulation structure for distributions of θ given background survey variables and design variables.

<u>Assumption 2</u>. **Item responses x are governed by the latent variable θ through a model of known functional form, $p(x|\theta,\beta)$, characterized by possibly unknown parameters β and satisfying the local independence properties (8) and (9). Independence is assumed over subjects, so that**

$$\begin{aligned} p(\underset{\sim}{x}|\underset{\sim}{\theta},\beta,\underset{\sim}{y},\underset{\sim}{Z}) &= p(\underset{\sim}{x}|\underset{\sim}{\theta},\beta) \\ &= \prod_{i}^{N} p(x_i|\theta_i,\beta)\ . \end{aligned} \qquad (12)$$

<u>Assumption 3</u>. **The distribution of latent variables given collateral survey variables y and design variables z follows a**

known functional form, $p(\theta|y,z,\alpha)$, **characterized by possibly unknown parameters** α. **Independence is assumed over subjects, so that**

$$p(\underset{\sim}{\theta}|\underset{\sim}{y},\underset{\sim}{Z},\alpha) = \prod_{i}^{N} p(\theta_i|y_i,z_i,\alpha) \ . \qquad (13)$$

It is important to note that these distributions are conditioned on the design variables z employed in the sampling design. While the sample design for selecting units from the existing population may be complex, sampling this population from the hypothetical superpopulation is SRS given z. This is the essence of the model-based approach to sampling-survey inference. It is used in this presentation not to handle uncertainty due to sampling examinees, but to build conditional distributions for latent variables, since it opens the door to the aforementioned methods of estimating the parameters of latent-variable distributions from SRS data.

To see how Assumptions 2 and 3 lead to $p(\underset{\sim}{\theta}|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{Z})$, first consider some relationships conditional on α and β. Using Bayes theorem, then (12) and (13), gives

$$p(\underset{\sim}{\theta}|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{Z},\alpha,\beta) = K_{\alpha\beta}\ p(\underset{\sim}{x}|\underset{\sim}{\theta},\underset{\sim}{y},\underset{\sim}{Z},\alpha,\beta)\ p(\underset{\sim}{\theta}|\underset{\sim}{y},\underset{\sim}{Z},\alpha,\beta)$$

$$= \prod_{i}^{N} K_{i\alpha\beta}\, p(x_i|\theta_i,\beta)\; p(\theta_i|y_i,z_i,\alpha) \;, \qquad (14)$$

where the normalizing constants $K_{i\alpha\beta} = p(x_i|y_i,z_i,\alpha,\beta)$ depend on α and β but not on θ_i. Given α and β, then, the predictive distribution for the latent variable θ_i of Subject i, or $p(\theta_i|x_i,y_i,z_i,\alpha,\beta)$, would be obtained by normalizing the product of (i) the likelihood function induced for θ by x_i via the latent-variable model (8), and (ii) the conditional distribution for θ implied by his or her background and design variables y_i and z_i.

Multiple Imputations for Latent Variables

The preceding paragraphs give the framework for randomization-based inference with latent variables in sample surveys. To operationalize the approach with multiple imputations requires specializing the procedure outlined above as follows:

1. Obtain the posterior distribution of the parameters β of the latent-variable model and α of the conditional distributions of θ, namely $p(\alpha,\beta|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{Z})$, by the methods discussed in connection with (10)--for example, a large sample normal approximation based on the MLE $(\hat{\alpha},\hat{\beta})$ and asymptotic covariance matrix $\Sigma_{\alpha\beta}$.

2. Produce M "completed" datasets $(\underset{\sim}{\theta}_{(m)},\underset{\sim}{x},\underset{\sim}{y})$. For the m^{th},

a. Draw a value $(\alpha,\beta)_{(m)}$ from $p(\alpha,\beta|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{Z})$. (If α and β have been estimated very precisely, it may be acceptable to use $(\hat{\alpha},\hat{\beta})$ for each completed dataset in what is commonly known as an empirical Bayes approximation. This expedient introduces a tendency to underestimate the uncertainty associated with final estimates of S.)

b. For each respondent, draw a value from the predictive distribution $p(\theta|x_i,y_i,z_i,(\alpha,\beta)=(\alpha,\beta)_{(m)})$. Taken together, the resulting imputation for θ_i and his or her observed values of x_i, y_i, and z_i constitute the "completed" response of Subject i.

3. Using each completed dataset, calculate $s_{(m)}=s(\underset{\sim}{\theta}_{(m)},\underset{\sim}{x},\underset{\sim}{y})$ and $U_{(m)}=U(\underset{\sim}{\theta}_{(m)},\underset{\sim}{x},\underset{\sim}{y})$.

4. The final estimate of S is the average of the M estimates from the completed datasets, or $s_M = \Sigma\, s_{(m)} / M$.

5. The estimated sampling variance of s_M as an estimate of S, namely V_M, is the sum of two components:

$$V_M = U_M + (1+M^{-1})\, B_M ,$$

where U_M and B_M are defined as for (5), to quantify uncertainty due to sampling subjects and the latency of θ, respectively.

6. If inferences about S would have been based on $(s-S) \sim N(0,U)$ had all responses been observed, inferences are now based on $(s_M-S) \sim t_\nu(0,V_M)$, a t-distribution with degrees of freedom ν given in (6).

These procedures apply to vector-valued latent variables θ as well as scalars. Extensions to vector-valued S are as discussed prcviously.

Steps 1 and 2 above produce M completed datasets that can be used to draw inferences about any number of sample statistics by applying Steps 3-6 repeatedly. An attractive feature of the approach is that the sophisticated methodologies and heavy computation are isolated in Steps 1 and 2, which can be carried out just once--probably by the institution held responsible for primary data analysis, where the necessary expertise and resources are more likely to be available. The completed datasets are then provided to secondary researchers, who need only apply standard routines for complete data M times and combine the results in simple ways.

Example 2: Multiple Imputation under Classical Test Theory

This example lays out imputation procedures for when the latent-variable model is the classical true-score test model with normal errors, and there are two collateral survey variables. In order to focus on the construction and the nature of imputations, it is assumed that a large SRS will be drawn from a multivariate normal population with known parameters. There are four variables for each subject:

θ, the latent variable, is examinee "true score;"

x is examinee observed score; and

y_1 and y_2 are collateral examinee variables.

Consistent with true-score test theory with normal errors, assume that $x = \theta + e$, where the residual or error term e is distributed $N(0,\sigma_e^2)$ independently of θ, y_1, and y_2. The latent variable model is thus

$$p(x|\theta) \sim N(\theta,\sigma_e^2) \ . \qquad (14)$$

Assume further that (θ,y_1,y_2) follows a standard multivariate normal distribution in the population, so that jointly, $(x,\theta,y_1,y_2) \sim MVN(\underset{\sim}{0},\underset{\sim}{\Sigma})$ with

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 1+\sigma_e^2 & & (sym) & \\ 1 & 1 & & \\ r_{\theta 1} & r_{\theta 1} & 1 & \\ r_{\theta 2} & r_{\theta 2} & r_{12} & 1 \end{bmatrix}$$

$\underset{\sim}{\Sigma}_y$ will refer to the covariance matrix for $y=(y_1,y_2)$.

The conditional distribution of θ given y is obtained as

$$p(\theta|y) \sim N(\beta' y,\sigma_{\theta|y}^2) \ , \qquad (15)$$

where

$$\beta' y \;=\; \beta_{1|2}\, y_1 + \beta_{2|1}\, y_2 \;=\; E(\theta|y), \text{ and}$$

$$\sigma^2_{\theta|y} = 1 - \beta' \,\Sigma_{\underset{\sim}{y}}\, \beta = 1 - R^2 ,$$

where R^2 is the proportion of variance of θ accounted for by y. In this example, $\beta_{1|2}=(r_{\theta 1}-r_{\theta 2}r_{12})/(1-r_{12}^2)$ and $\beta_{2|1}=(r_{\theta 2}-r_{\theta 1}r_{12})/(1-r_{12}^2)$. It follows that $p(x|y) \sim N(\beta' y, \sigma^2_{x|y})$, with $\sigma^2_{x|y} = \sigma^2_{\theta|y} + \sigma^2_e$.

In addition to the multiple regression coefficients there are simple regression coefficients for θ or x on y_1 or y_2 alone; e.g.,

$$E(\theta|y_1) = E(x|y_1) = \beta_1\, y_1 = (\beta_{1|2} + r_{12}\, \beta_{2|1})\, y_1 .$$

In this example, $\beta_1=r_{\theta 1}$ and $\beta_2=r_{\theta 2}$.

Define the "conditional reliability" ρ_c of x as a measure of θ given y as

$$\rho_c \;=\; \sigma^2_{\theta|y} \,/\, (\sigma^2_{\theta|y} + \sigma^2_e) .$$

Note that $0 \le \rho_c \le 1$ and $\rho_c \sigma^2_{x|y} = \sigma^2_{\theta|y}$.

An imputation for a sampled subject will be drawn from a predictive distribution of the form $p(\theta|x,y)$ which is proportional to $p(x|\theta)\, p(\theta|y)$. The first factor in this product is the likelihood, which in this example is $N(x,\sigma^2_e)$; the MLE is in fact simply x. The second factor is the conditional density of θ given y, which is $N(\beta' y, \sigma^2_{\theta|y})$. By Kelley's (1947) formula,

$$p(\theta|x,y) \sim N(\bar{\theta},\sigma^2_{\theta|xy}) ,$$

where

$$\bar{\theta} = E(\theta|x,y) = \rho_c x + (1-\rho_c)\ \beta' y$$

and

$$\sigma^2_{\theta|xy} = Var(\theta|x,y) = (1-\rho_c)\ \sigma^2_{\theta|y} = (1-\rho_c)(1-R^2) .$$

An imputation $\tilde{\theta} = \tilde{\theta}(x,y)$ is thus constructed as

$$\tilde{\theta} = \bar{\theta} + f ,$$

where f is drawn at random from $N(0,\sigma^2_{\theta|xy})$.

For a given individual, an imputation is not an optimal point estimate of θ. It is neither unbiased nor efficient, as is the MLE $\hat{\theta}$=x; nor does it minimize mean square error over the population, as does $\bar{\theta}$. But it can be shown that the distribution of $(\tilde{\theta},y)$ is multivariate normal with the same mean vector and covariance matrix as that of (θ,y). For the population mean, for example,

$$E(\tilde{\theta}) = E[\rho_c x + (1-\rho_c)\ \beta' y + f]$$

$$= \rho_c E(x) + (1-\rho_c)\ \beta'\ E(y) + E(f)$$

$$= \rho_c\ 0 + (1-\rho_c)\ \beta'\ \underset{\sim}{0}\ + 0$$

$$= E(\theta) \ .$$

For the covariance between $\tilde{\theta}$ and y_1,

$$\begin{aligned}
\mathrm{Cov}(\tilde{\theta},y_1) &= E\{[\rho_c x + (1-\rho_c)\ \beta' y + f]\ y_1\ \} \\
&= E[\rho_c x y_1] + E[(1-\rho_c)(\beta_{1|2} y_1 + \beta_{2|1} y_2)\ y_1] + E[f y_1] \\
&= \rho_c\ r_{\theta 1} + (1-\rho_c)[\beta_{1|2}\ E(y_1{}^2) + \beta_{2|1}\ E(y_1 y_2)] + 0 \\
&= \rho_c\ r_{\theta 1} + (1-\rho_c)\ (\beta_{1|2} + \beta_{2|1}\ r_{12}) \\
&= \rho_c\ r_{\theta 1} + (1-\rho_c)\ r_{\theta 1} \\
&= r_{\theta 1} \\
&= \mathrm{Cov}(\theta,y_1) \ .
\end{aligned}$$

By similar calculation,

$$\begin{aligned}
\mathrm{Var}(\tilde{\theta}) &= 1 && = \mathrm{Var}(\theta) \ , \\
E(\tilde{\theta}|y) &= \beta' y && = E(\theta|y) \ , \\
\mathrm{Var}(\tilde{\theta}|y) &= \sigma^2_{\theta|y} && = \mathrm{Var}(\theta|y) \ , \\
E(\tilde{\theta}|y_1) &= \beta_1\ y_1 && = E(\theta|y_1) \ , \text{ and} \\
\mathrm{Var}(\tilde{\theta}|y_1) &= 1 - r^2_{\theta 1} && = \mathrm{Var}(\theta|y_1) \ .
\end{aligned}$$

Given that the mean of $\tilde{\theta}$ is an unbiased estimate of the mean of θ, how much does uncertainty increase when $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are observed rather than $\underset{\sim}{\theta}$ itself? For an SRS of size N, the SEM^2 for N observations of θ is U=1/N. For large M, the variance among estimates of the population mean from the completed datasets is the variance of the average of N realizations of f, or $\sigma^2_{\theta|xy}/N$. The total variance for an estimate of the mean based on many imputations is the sum of these components, representing an inflation from 1/N of $(1-\rho_c)(1-R^2)$ percent--the proportion of variance in θ unexplained by x and y.

It has been shown that treating $\tilde{\theta}$s as θs, one obtains the correct expectations for estimates of population attributes such as percentile points, conditional distributions, and proportions-of-variance-accounted-for. In contrast, treating either $\hat{\theta}$s or $\bar{\theta}$s as θs one obtains estimates for some attributes that have the correct expectations for some attributes, but not for others. For estimating the mean of θ, all three turn out to have the correct expectation of zero. For the variance, the expectation using imputations is the correct value of one, but the expected variances of $\hat{\theta}$ and $\tilde{\theta}$ are $1+\sigma^2_e$ and $1-\sigma^2_{\theta|xy}$ respectively.

Table 1 gives the expectations of estimates of some population attributes that result when various point estimates of θ are treated as if they were θ. Imputations $\tilde{\theta}$ appear there, as well as the MLE $\hat{\theta}$ (=x). The Bayes estimate referred to above as $\bar{\theta}$ is denoted $\bar{\theta}_{xy}$ to indicate that it is conditional on both x <u>and</u> y; another Bayes estimate often used in practice, denoted in the table as $\bar{\theta}_x$, ignores y. The variance of $\bar{\theta}_x$ is known to be less than that of θ, so it is sometimes suggested that $\tilde{\theta}_x$ values be inflated by the factor needed to bring their variance up to that of θ. The resulting rescaled Bayes estimate is denoted by $\tilde{\theta}_x^{(r)}$. While this rescaling corrects the variance of the population as a whole, it does not completely remove the biases associated with attributes of conditional distributions involving y.

==================

Table 1 about here

==================

Note that the distortions in secondary analyses of all these point estimates depend on test reliability. Reliability, and therefore the magnitudes of biases, will vary if test lengths differ over time or across respondent groups. Tables 2 and 3 illustrate this point by giving numerical values for the expressions in Table 1 that are obtained from a test with ρ_T=.50 and a test with ρ_T=.91, in both cases with $r_{\theta 1} = r_{\theta 2} = r_{12} = .50$. The first test has a reliability that might be expected with ten items on a particular topic that appear on an educational assessment instrument. The second has a reliability more like that of a 60-item achievement test. The biases that occur when using any of the "optimal" point estimates instead of the imputations, are readily apparent for the short test but are less serious for the long test.

Tables 2 and 3 here

Along with moments of population and subpopulation distributions, cumulative probabilities are sometimes required--e.g., $P(\theta \geq 1)$ or $P(\theta \geq 1 | y_1=-1)$. Statistics of this type are important in NAEP, where selected points along the proficiency are anchored by behavioral descriptions, and the proportions of populations and subpopulations above these points are tracked over time (NAEP, 1985). $P(\theta \geq 1)$, in this example, is the proportion of the total population above one standard deviation in the standard normal distribution, or .1587. $P(\theta \geq 1 | y_1=-1)$ is the proportion of the subpopulation defined by $y_1=-1$ with θ-values higher than one standard deviation above the total population mean, which, in this example, is .0418, the proportion of $N(-.5,\sqrt{.75})$ above 1. Table 4 gives the expectations of these values that obtain when point estimates of θ are treated as θ. Even though the regression estimates for β_1 from $\hat{\theta}$ and $\bar{\theta}_{xy}$ are unbiased, the estimated population proportions above the cut point are distorted. Again the distortion is less serious with the long test.

Insert Table 4 here

Computing Approximations and Secondary Biases

Obtaining consistent estimates of population attributes with multiple imputations requires drawing imputations from consistent estimates of the correct predictive distributions $p(\theta|x,y)$. Assuming the latent variable model $p(x|\theta)$ is specified correctly, it is possible to obtain detailed nonparametric approximations of $p(\theta|y)$, and therefore of $p(\theta|x,y)$, when the dimensionalities of θ and y are low--say less than five latent variables and five collateral variables (Mislevy, 1984). When the dimensionalities of θ and y are high, however, as in NAEP with its hundreds of background and attitude items, simplifications and computing approximations cannot be avoided. This section lays out a general framework for the problems entailed by using simplified approximations of $p(\theta|y)$, derives some explicit results for the true-score example introduced above, and offers guidelines for practical applications.

The Nature of the Problem

The imputation-based estimate $s_M(\underset{\sim}{x},\underset{\sim}{y})$ approximates the expectation of $s(\underset{\sim}{\theta},\underset{\sim}{x},\underset{\sim}{y})$ defined in (11) by evaluating s with draws $\underset{\sim}{\tilde{\theta}}$ from

$$p(\underset{\sim}{\theta}|\underset{\sim}{x},\underset{\sim}{y}) \propto$$

$$\iint p(\underset{\sim}{x}|\underset{\sim}{\theta},\beta)\ p(\underset{\sim}{\theta}|\underset{\sim}{y},\underset{\sim}{z},\alpha)\ p(\alpha,\beta|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{z})\ d\alpha\ d\beta\ . \qquad (16)$$

As noted above, however, procedures for characterizing $p(\underset{\sim}{\theta}|\underset{\sim}{y},\underset{\sim}{z},\alpha)$ are neither numerous nor easily applied. It becomes necessary to approximate these conditional distributions by some tractable form, such as multivariate normal with a common residual variance (Mislevy, 1985). Rather than conditioning on all elements of y and z and their interactions, which may number into the millions, approximations based on perhaps fewer than a hundred effects must suffice. The correct conditional distribution is replaced by the computing approximation $p^*(\theta|y,z,\alpha^*)$, which, when combined with the latent variable model as in (16), yields the computing approximation from which *imputations* $\underset{\sim}{\tilde{\theta}}^*$ are drawn:

$$p^*(\underset{\sim}{\theta}|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{z}) \propto$$

$$\iint p(\underset{\sim}{x}|\underset{\sim}{\theta},\beta)\ p^*(\underset{\sim}{\theta}|\underset{\sim}{y},\underset{\sim}{z},\alpha^*)\ p(\alpha^*,\beta|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{z})\ d\alpha^*\ d\beta\ . \qquad (17)$$

While the expectation of s_M based on imputations $\underset{\sim}{\tilde{\theta}}$ has the correct value s, the expectation of s_M^* based on imputations $\underset{\sim}{\tilde{\theta}}^*$ may not. Its expectation is

$$E(s_M^*) = \int s(\underset{\sim}{\theta},\underset{\sim}{x},\underset{\sim}{y})\ p^*(\underset{\sim}{\theta}|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{z})\ d\underset{\sim}{\theta}\ . \qquad (18)$$

For a fixed observed sample $(\underset{\sim}{x},\underset{\sim}{y})$, the bias in estimating s caused by using $p^*(\theta|y,z,\alpha^*)$ rather than $p(\theta|y,z,\alpha)$ is thus

$$\text{Bias} = E(s_M^*|\underset{\sim}{x},\underset{\sim}{y}) - E(s|\underset{\sim}{x},\underset{\sim}{y})$$

$$= \int s(\underset{\sim}{\theta},\underset{\sim}{x},\underset{\sim}{y})\ [p^*(\underset{\sim}{\theta}|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{z})-p(\underset{\sim}{\theta}|\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{z})]\ d\underset{\sim}{\theta}\ . \qquad (19)$$

Rubin raises the possibility of biases in secondary analyses when he mentions that the imputer's model may differ from the analyst's (Rubin, 1987, pp. 107, 109-112, 151). His analyses suggest that biases for population means and variances may be mild when filling in a modest number of missing responses in standard surveys, particularly if the imputer's model is more inclusive than the analyst's. It will become apparent that thic conclusion need not hold in the context of latent variables models, however, especially when practical constraints force the imputer to use a less inclusive model than the analyst.

Example 2 (continued)

Consider again the multinormal example introduced above, with its true-score latent variable model $x|\theta \sim N(\theta,\sigma_e^2)$ and population model $\theta|y \sim N(\beta'y,\sigma_{\theta|y}^2)$. Suppose now that the imputer conditions on y_1 but not y_2 when building the model from which imputations are drawn. That is, the correct predictive distribution for imputations is

$$p(\theta|x,y) = N[\rho_c x + (1-\rho_c)\beta'y,\ \sigma_{\theta|xy}^2] ,$$

but the imputer draws from

$$p^*(\theta|x,y) = N[\rho_c^* x + (1-\rho_c^*)\beta_1 y_1,\ \sigma_{\theta|xy1}^2] ,$$

where ρ_c^* is the conditional reliability of x given y_1 but not y_2, or

$$\rho_c^* = \sigma^2_{\theta|y1} / (\sigma^2_{\theta|y1} + \sigma^2_e) = (1-r^2_{\theta 1}) / (1-r^2_{\theta 1} + \sigma^2_e) ;$$

and

$$\sigma^2_{\theta|xy1} = \mathrm{Var}(\theta|x,y_1) = (1-\rho_c^*)\, \sigma^2_{\theta|y1} = (1-\rho_c^*)(1-r^2_{\theta 1}) .$$

An imputation for Subject i now takes the form

$$\tilde{\theta}_i^* = \rho_c^* x_i + (1-\rho_c^*)\beta_1 y_{i1} + g ,$$

where g is a draw from $N(0,\sigma^2_{\theta|xy1})$.

How do the attributes of the distribution of imputed values $\tilde{\theta}^*$ fare as estimates of the attributes of the distribution of θ? By calculations similar to those in the first part of the example, it can be shown that $(\tilde{\theta}^*,y)$ is normal with mean vector $\underset{\sim}{0}$, as is (θ,y), and its covariance matrix agrees with that of (θ,y) for all elements except the one for $\tilde{\theta}^*$ with y_2. Rather than $r_{\theta 2}$, one obtains

$$\mathrm{Cov}(\tilde{\theta}^*,y_2) = r^*_{\theta 2} = \rho_c^*\, r_{\theta 2} + (1-\rho_c^*)\, r_{\theta 1}\, r_{12} .$$

It follows first that characteristics of the joint distribution of $\tilde{\theta}^*$ and y_1 are identical to those of θ and y_1. For instance,

$$E(\tilde{\theta}^*) = 0 = E(\theta)$$

and

$$\mathrm{Var}(\tilde{\theta}^{*}) = 1 = \mathrm{Var}(\theta) .$$

This corresponds to the case in which the imputer's model is more inclusive than the analyst's, and, as Rubin suggests, the result is satisfactory. Also, corresponding to the case in which the imputer and the analyst use the same model, one obtains

$$E(\tilde{\theta}^{*}|y_1) = \beta_1 y_1 = E(\theta|y_1)$$

and

$$\mathrm{Var}(\tilde{\theta}^{*}|y_1) = 1-\sigma^2_{\theta|xy1} = \mathrm{Var}(\theta|y_1) .$$

The secondary analyst thus obtains the correct expectations for both marginal analyses of θ and conditional analyses involving only θ and the collateral variable that was conditioned on.

Less salubrious are results for analyses involving y_2, the collateral variable that was not conditioned on. Whereas

$$E(\theta|y) = E(\tilde{\theta}|y) = \beta_{1|2}y_1 + \beta_{2|1}y_2 ,$$

one finds that

$$E(\tilde{\theta}^{*}|y) = [\beta_{1|2} + (1-\rho_c^{*})\beta_{2|1}r_{12}]y_1 + \rho_c^{*}\beta_{2|1}y_2$$

$$= \beta_{1|2}y_1 + \beta_{2|1}[\rho_c^{*}y_2 + (1-\rho_c^{*})E(y_2|y_1)] \qquad (20a)$$

$$= \beta_{1|2}y_1 + \beta_{2|1}y_2 - (1-\rho_c^{*})\beta_{2|1}(y_2-r_{12}y_1) . \qquad (20b)$$

A bias is thus introduced into the imputations, the nature of which is to attenuate the contribution from y_2, the omitted variable. Equation (20a) shows that the contribution associated with y_2 is a weighted average of (i) the correct contribution--$\beta_{2|1}y_2$--to the degree that x is a reliable indicator of θ, and (ii) to the degree that x is unreliable, a contribution associated with the expectation of the omitted variable given the observed value of the conditioned variable. Equation (20b) shows that the bias can be driven to zero in three ways:

a. As $\rho^*_c \to 1$; i.e., x is a perfectly reliable measure of θ;

b. As $\beta_{2|1} \to 0$; i.e., there is no contribution from y_2 anyway;

c. As $r_{12}y_1 \to y_2$ for all y_1; i.e., y_2 is perfectly predictable from y_1.

Some consequences for regression analyses involving y_2 are now considered.

Whereas the regression coefficient for y_2 in the multiple regression of θ on y is $\beta_{2|1}$, the corresponding coefficient for y_2 in the multiple regression of $\tilde{\theta}^*$ on y is

$$\beta^*_{2|1} = \rho^*_c \beta_{2|1} . \tag{21}$$

The expected regression coefficient has been shrunken by the factor $(1-\rho^*_c)$, the complement of test reliability given y_1.

Whereas the regression coefficient for y_1 in the multiple regression of θ on y is obtained as

$$\beta_{1|2} = (r_{\theta 1}-r_{\theta 2}r_{12})/(1-r_{12}^{2}) \ ,$$

the corresponding coefficient in the multiple regression of $\tilde{\theta}^*$ on y is

$$\beta^*_{1|2} = (r_{\theta 1}-r^*_{\theta 2}r_{12})/(1-r_{12}^{2}) \ . \qquad (22)$$

The bias can be expressed as

$$\beta^*_{1|2} - \beta_{1|2} = (1-\rho^*_c)\ \beta_{2|1}\ r_{12} \ .$$

Thus, in the analysis of conditional θ distributions given both y_1 and y_2, bias exists for the coefficient of y_1 even though it has been conditioned on. Since $r_{12}y_1 = E(Y_2|y_1)$, the character of the bias is to absorb a portion of the unique contribution of the nonconditioned variable, to the extent that x is unreliable.

Whereas the coefficient β_2 for the simple regression of θ on y_2 is $r_{\theta 2}$, the corresponding coefficient for $\tilde{\theta}^*$ on y_2 is

$$\beta^*_2 = r^*_{\theta 2} = \rho^*_c\ r_{\theta 2} + (1-\rho^*_c)\ r_{\theta 1}\ r_{12} \ . \qquad (23)$$

The bias with which β_2 is estimated from $\tilde{\theta}^*$ can be expressed as

$$\beta^*_2 - \beta_2 = (1-\rho^*_c)\ (1-r_{12}^{2})\ \beta_{2|1} \ .$$

This bias is reduced as either ρ^*_c, the test reliability, or r^2_{12}, the proportion of y_2 predictable by y_1, approaches one, or as the unique contribution of y_2 in predicting θ approaches zero.

To summarize, the degree of biases in secondary analyses involving variables not conditioned on involves (i) the reliability of x, (ii) the association between the conditioned and the nonconditioned collateral variables, and (iii) the unique contribution of the nonconditioned variable to predicting θ. Higher values of ρ^* are unequivocally helpful, as they reduce biases of all types. Higher values of r_{12}^2, on the other hand, <u>mitigate</u> bias in simple regression involving only the nonconditioned variable, but <u>exacerbate</u> the bias for the coefficient of the conditioned variable in multiple regression involving both. These conclusions extend in natural ways to sets of conditioned and nonconditioned variables (Beaton and Johnson, 1987; Mislevy and Sheehan, 1987).

Table 5 gives values for expected regression coefficients in analyses of θ and $\tilde{\theta}^*$, using the numerical values employed in Tables 2 and 3. For simple regression, the results for the conditioned variable are unbiased, and the results for the nonconditioned variable are comparable in accuracy to those obtainable from "optimal" point estimates (bearing in mind the fact that MLEs yield unbiased regression coefficients but biased conditional variances and percentile points). It can also be seen that the results of multiple regression are more sensitive to omitting variables than those of simple regression. These findings underscore the importance of choosing conditioning variables wisely if practical considerations preclude conditioning on all the background variables that have been surveyed.

==================

Insert Table 5 here

==================

Implications and Recommendations

The foregoing results indicate that care is required to impute values for latent variables in a way that leads to acceptably accurate results in secondary analyses. Precise determination of θ by item responses x alleviates the problem of biases, but to do so in the context of educational or psychological measurement requires large numbers of item responses from every subject--a design that is inefficient for drawing inferences about only population attributes. When testing time for individuals is limited, the imputer must build a computing approximation $p^*(\theta|y,z)$ that gives good results for a broad range of potential statistics s involving $\underset{\sim}{\theta}$.

If there are only a small number of values that y and z can take, and a large number of subjects at each combination of values, one can obtain a nonparametric estimate for each (y,z) combination by the methods of Laird (1978) or Mislevy (1984). This leads to imputations that are free from specification error in $p(\theta|y,z)$, and secondary analyses will not suffer from biases from this source. This approach is simply not possible, though, for surveys such as NAEP with large numbers of background and attitude items.

As mentioned above, one possibility is to assume normal conditional distributions with structured means and a common residual variance (Mislevy, 1985). In addition to assuming this tractable distributional form, further simplification becomes necessary when there are large numbers of design and background variables. In ANOVA terms, conditioning on the full joint distribution of (y,z) could involve millions of effects, while currently available computing procedures can handle up to about a hundred. Specifying p^* wisely means choosing contrasts so as to optimize the accuracy of potential secondary analyses. Based on the results of Example 2, the following advice can be offered:

Determine θ as well as is practical. In the context of measuring latent proficiencies by test items, recall the decreasing rate at which reliability increases--and potential biases in secondary analyses thereby decrease--with additional test items. Trade-offs arise among potential item-sampling designs. Compared to a design that gives five items to each subject, a design that gives ten items yields less efficient estimates of statistics involving variables y^c that have been conditioned on, but less biased estimates of those involving variables y^{nc} that have not been conditioned on.

Borrow information from related scales. As noted earlier, imputation methods apply to vector-valued latent variables. In such applications, one estimates multivariate conditional distributions $p(\theta|y,z)$, combines them with multidimensional likelihoods $p(x|\theta)$, and draws vector-valued imputations from joint predictive distributions $p(\theta|x,y,z)$. This was done in the NAEP survey of Young Adult Literacy (Kirsch and Jungeblut, 1986), where each respondent was presented five to fifteen items in each of four IRT literacy scales. The population correlations of about .6 among scales sharpened the predictive distribution of each scale for an individual: while the information available directly about the scale was worth ten items, the thirty items from the other scales indirectly contributed information worth about another ten. The biases in secondary analyses were thus reduced as much by using multivariate imputations for the four scales jointly as they would have been under separate univariate imputations with twice as many items.

Condition explicitly on contrasts that are particularly important, such as treatment group in a survey designed expressly to compare treatment effects in a program evaluation. By doing so, one ensures that the marginal subpopulation means or regression coefficients involving key variables are estimated as accurately as possible. Note that y^c can include interactions as well as main effects.

Condition on well-chosen combinations of variables. Given current computational limits, it will often be impossible to condition on the main effects of all background variables, let alone two-way or higher interactions among them. One can reduce biases for a large number of contrasts of interest, beyond those that can be conditioned on explicitly, by conditioning on linear combinations of contrasts--for example, the first h component scores from a principal components decomposition of the covariance matrix among effects. The results of Example 2 imply that the variation these partially-conditioned-upon variables share with the explicitly-conditioned-upon component scores will have salutary effects in secondary analyses. The degree of bias for an effect will be limited to the proportion of its variance unaccounted for by the conditioned-upon components, times the complement of conditional test reliability.

Example 3: The 1984 NAEP Reading Assessment

During the 1983-84 school year, the National Assessment of Educational Progress (NAEP) surveyed the reading and writing skills of national probability samples of students at ages 9, 13, and 17, and in the modal grades associated with those ages, namely 4, 8, and 11. Beaton (1987) gives details of assessment procedures and analyses. This section of the present paper highlights the multiple-imputations procedures used in the analysis of the reading data.

The Student-Sampling Design

A multistage probability sampling design was employed to select students into the NAEP sample, with counties or groups of counties as the primary sampling units (PSUs). Schools served as second-stage sampling units. The assignment of testing sessions of different types to sampled schools was the third stage of sampling, and the selection of students within schools was the fourth. A total of 64 PSUs appeared in the sample, and assessments were administered at 1,465 schools. About 20,000 students were assessed in reading at each grade/age cohort. For convenience, grade/*age cohorts* will be referred to below simply by their age designations.

Sampling was stratified at the first stage according to geographic regions, Census Bureau "Sample Description of Community" (SDOC) classes, and, within urban and rural SDOC classes, a measure of SES; the latter two criteria comprise Size and Type of Community (STOC) classes. Selection probabilities of sampling units were inversely proportional to estimated population size, except that extreme rural and low-SES urban areas were oversampled by a factor of two. Neglecting minor adjustments for nonresponse and poststratification, the design variables Z were therefore region, STOC, PSU, and school membership.

Population means and totals of survey variables were computed as weighted sample means and totals, with a student's weight essentially inversely proportional to his or her probability of selection. Uncertainty of such a statistic s due to student sampling was approximated with a multiweight jackknife procedure. Thirty-two pairs of similar PSUs were designated. Approximating the uncertainty of s required computing it 33 times: once in a run with the total sample, and once with a run corresponding to each PSU pair, with one of its members left out of the sample but with the sampling weight of its partner doubled. The variance of the 32 jackknife estimates around the total value is U, the estimated sampling variance of s around S.

The Survey Variables

Each student responded to a number of survey items (Y) tapping demographic status, educational background and reading practices, and attitudes about reading and writing. About 50 were common to all assessment forms. Examples are gender, parents' education, ethnicity, and time spent watching television. Another 300, of which a given student would receive between about 10 and 30 under the assessment's balanced incomplete block (BIB) item-sampling design, addressed reading activities in the home and school.

A total of 340 multiple-choice and free-response reading exercises were used in the assessment, although a student who received any reading exercises received between 5 and 50 of them under the BIB design. About 80 percent of the students received some reading exercises. A few of the exercises appeared at all three ages, but most appeared in only one or in two adjacent ages.

The Latent-Variable Model

A priori considerations and extensive dimensionality analyses supported summarizing responses to a subset of 228 of the 340 items by an IRT model for a single underlying proficiency variable θ (Zwick, 1987). Responses to these items will be denoted by x. The 3-parameter logistic (3PL) IRT model was used. Under the 3PL, the probability of a correct response to Item j from a student with proficiency θ is given by

$$P(x_{(j)}=1|\theta,a_j,b_j,c_j) = c_j + (1-c_j)/\{1+\exp[-1.7a_j(\theta-b_j)]\} ,$$

where x=1 indicates a correct response and x=0 an incorrect one, while (a_j,b_j,c_j) are parameters that characterize the regression of $x_{(j)}$ on θ. Coupled with this model for a single item, the assumptions of local independence embodied by (8) and (9) give the likelihood function $p(x|\theta,\beta)$ $[=p(x|\theta,\beta,y,z)]$ for a response vector x to any subset of the 228 items. Here β denotes the vector of item parameters of all 228 items in the scale.

Students receiving any of the 228 reading items in the scale at all received between 5 and 40 of them, with the average about 17. The responses of a sample of 10,000 students were used with a modified version of Mislevy and Bock's (1983) BILOG computer program to estimate 3PL item parameters. The modifications allowed the c-parameters of free-response items to be set to zero a priori, and distinguished age cohorts when computing marginal probabilities of students' response patterns. The latter extension is necessary to achieve consistent item parameter estimates since items were assigned to cohorts based on prior knowledge about the proficiencies of the cohorts--younger students were generally administered easier items--but that consistency holds whether or not sampling weights within cohorts are used (Mislevy and Sheehan, in press). The resulting item parameter estimates were placed on a scale in which the unweighted calibration sample of students was standardized. This scale is arbitrary up to a linear transformation, so any other convention would have served equally well to set the scale. The resulting estimates were taken as fixed throughout the remainder of the study, thereby fixing the origin and unit-size of the θ scale.

The Imputation Model

The ideal population model $p(\theta|y,z)$ to use in conjunction with $p(\theta|x,\beta)$ is a joint distribution involving hundreds of survey variables and background and attitude items. The computer program available to estimate a latent population distribution was a prototype developed for Mislevy's (1985) 4-group example, with ANOVA-type structures on the means and normal residuals with a common within-group variance. In the time available to meet NAEP reporting deadlines, it was possible to extend the program to a design with 17 effects. A main effects model was chosen that focused accuracy on traditional NAEP reporting categories: sex, ethnicity, STOC, region, and parental education, along with indicators for at, above, or below modal grade and age. A "miscellaneous" cell was included in the model for the small fraction of students whose values on the aforementioned items were unrecoverable. Altogether, these variables comprised the vector $(y,z)^c$. The following model was assumed for conditional distributions:

$$\theta | y,z \sim N[(y,z)^{c}{}'\Gamma, \sigma^2] ,$$

where Γ is a vector of seventeen main effect parameters and σ^2 is the residual variance. Together, Γ and σ^2 comprise the superpopulation parameter α referred to in preceding sections.

Note that the design variables Z are captured largely but not completely, as STOC and region have been included but PSU, school membership, and interaction terms have not been. Because PSU and school-level variation are largely explained by the conditioned-upon region and SES indicators STOC and parental education, however, biases caused the omission of PSU and school membership in $(y,z)^c$ are largely mitigated.

Maximum likelihood estimates of Γ and σ^2 were obtained separately within ages using all available reading data--over 20,000 respondents per age. Separate age analyses implicitly allow for all two-way interactions between cohort and each of the other effects. The results are shown in Table 6. Like item parameter estimates, these estimates were also taken as known true values thereafter.

==================

Table 6 about here

==================

The posterior distribution of each student i was approximated by a histogram over 40 equally spaced points Θ between -4.785 and +4.785. The weight of the q^{th} bar in the histogram for the i^{th} student was obtained as follows:

$$P(\Theta_q | x_i, y_i^c, z_i^c) \sim \frac{P(x_i | \theta=\Theta_q, \beta=\hat{\beta}) \; P(\theta=\Theta_q | y_i^c, z_i^c, \alpha=\hat{\alpha})}{\sum_s P(x_i | \theta=\Theta_s, \beta=\hat{\beta}) \; P(\theta=\Theta_s | y_i^c, z_i^c, \alpha=\hat{\alpha})} , \qquad (24)$$

where $P(\theta=\Theta_s | y_i^c, z_i^c, \alpha=\hat{\alpha})$ is the density at Θ_s of the normal pdf with mean $(y,z)_i^{c\prime}\hat{\Gamma}$ and the residual variance for the age group to which Student i belongs. Each imputation was drawn from this distribution in two steps. First, a bar was selected at random in accordance with the weights determined in (24). Second, a point was selected at random from a uniform density over the θ range spanned by the selected bar. (Logistic interpolation would have been better.) Five imputations were drawn in this manner for each student in the sample.

Illustrative Results

The results of primary interest in this first analysis of the 1984 data were the mean proficiencies of the subpopulations determined by the traditional NAEP reporting categories. A given weighted mean was calculated five times, once from each completed data set. The average of the five was the reported estimate. A jackknife variance was also calculated from each completed dataset, with the imputations in the set treated as known true values of θ; these values were also averaged, to estimate U. The reported variance was the sum of this average jackknife variance and $1+M^{-1}=1.2$ times the variance of the five aforementioned means. In order to avoid negative numbers and fractional values, thc original θ scale was linearly transformed by $50\theta + 250.5$. Table 7 illustrates some results for Age 17. The full public report is available as The Reading Report Card: Progress toward Excellence in our Schools (NAEP, 1985).

==================

Table 7 about here

==================

The proportional increase in variance due to the latency of θ, denoted earlier as r_M, varies from 2 percent up to nearly 30 percent. The largest increase is associated with a lower-than-average scoring group, for whom the test items were relatively more difficult. The proficiencies of low-scoring individuals were determined less precisely by their item responses, so that the likelihoods induced for θ, and the consequent posterior distributions $p(\theta|x,y,z)$ from which their imputations were drawn, were more dispersed. Estimated means from such subpopulations tended to vary more widely across completed datasets than would means for groups of similar size whose typical members were measured more accurately by their item responses.

Biases in Secondary Analyses

The completed datasets described above were constructed so as to focus their accuracy on the marginal subpopulation means featured in The Reading Report Card. As discussed in a previous section, however, analyses involving survey variables that were not conditioned on are subject to bias. An opportunity soon arose to examine such biases. A report on reading proficiency levels among pupils whose primary language was not English came due, the analysis plan for which specified multiple regression analyses involving both variables that had and variables that had not been conditioned on when imputations for the original analysis were constructed. Aware that the proposed analyses were sensitive to failure to condition, the NAEP staff created new completed datasets in which all the variables required in the analyses were conditioned upon. This was made possible by Sheehan's (1985) M-GROUP computer program for estimating Γ and σ^2 in larger models. Some fifty effects were included in the recalculations for each age. The same multiple regression analyses were carried out twice, once with the original completed datasets and once with new ones created with extended conditioning vectors. The results of one such comparison are summarized in Table 8. Baratz-Snowden and Duran (1987) give the final results of all runs.

=================

Table 8 about here

=================

It may be seen that multiple regression coefficients for the two effects which were conditioned upon in both analyses were least affected in the recalculation, differing by amounts only 4 and 9 percent of their new estimated values. Coefficients for significant effects not originally conditioned on, however, differed by between 10 and 40 percent. The direction of the difference, in nearly all of these cases, was that the original estimates were shrunken toward zero. Performing the analysis on the original completed datasets would correctly inform the researcher about the directions of effects, but would tend to underestimate their magnitudes by an average of 30-percent.

Extensions

The experience with multiple imputation procedures gained with the 1984 reading assessment led to a number of insights on how to extend or improve the procedures. Four are mentioned below.

Multivariable Imputation. The preceding discussions have concentrated on the case of a single latent variable. While this proved adequate for summarizing reading data, both empirical and theoretical evidence demonstrate the need for multiple scales in broader content areas such as mathematics and science. NAEP extended multiple imputations procedures to the case of four variables in the Young Adult Literacy Assessment (Kirsch and Jungeblut, 1986), and later applied these procedures to five subscales each to analyze the 1986 mathematics and science assessments (Beaton, 1988).

In the multivariable case, each latent variable--a different aspect of, say, literacy skill--is defined through an IRT model. Assuming conditional independence, the four-dimensional likelihood $p(x_1,...,x_4|\theta_1,...,\theta_4)$ is simply the product of the four univariate IRT likelihoods, or $\Pi\ p(x_k|\theta_k)$. The conditional distributions $p(\theta_1,...,\theta_4|y,z)$ are generally <u>not</u> independent, however, their associations reflecting population correlations among skills. The predictive distributions $p(\theta_1,...,\theta_4|x_1,...,x_4,y,z)$ from which imputations are drawn reflect these associations. Compared to carrying out imputation procedures separately within each scale, the multivariable solution exploits information from all scales to strengthen inferences about each, and yields consistent, rather than attenuated, estimates of association among the scales.

<u>Conditioning on Principal Components</u>. An aspect of multiple imputation procedures that requires improvement is the accuracy of multiple regression analyses that include nonconditioned background variables. Conditioning on <u>more</u> background variables with Sheehan's (1985) improved M-GROUP program certainly increases the number of secondary analyses whose accuracy will be improved,

but research is currently under way to determine how to choose them wisely. As mentioned above, conditioning on well-chosen linear combinations of large numbers of variables holds promise. Analyses of the 1988 data will examine gains in accuracy attainable with different combinations of numbers of effects conditioned upon explicitly and effects conditioned on partially, though principal components.

Accounting for Uncertainty in α and β. Analyses to date have taken estimates of item parameters β and conditional distribution parameters α as known, thereby neglecting the component of uncertainty associated with them. Present indications are that the resulting overstatement of precision is negligible because of the huge sample sizes from which these effects are estimated, but research is under way to develop efficient methods for incorporating uncertainty about them as well as about θ. One approach to doing so is leans on asymptotic results, drawing from multivariate normal (α,β) distributions with means given by MLEs and variances given by inverses of information matrices. An alternative approach is to draw from multivariate distributions whose mean and variance matrix were obtained by a jackknife procedure. The latter is more intensive computationally, but captures variation due to lack of model fit as well as due to sampling.

The Average Response Method (ARM). To analyze the data from the 1984 NAEP survey of writing, Beaton and Johnson (1987) worked out multiple imputation procedures for the setting of general linear models. They address the problem of characterizing the distribution of the average of ratings over all writing exercises--a straightforward problem when every examinee is presented all exercises, but effectively a latent variable problem under an item-sampling design in which each examinee takes only a few exercises from the pool. Computational procedures are simpler under the ARM than with IRT models because the assumed linearity of relationships permits noniterative unweighted least squares solutions. Expressions for estimation, imputation, and expectation in secondary analyses offer insight into the problem for those familiar with the theory of general linear models.

Conclusion

At the beginning of the decade, Bock, Mislevy, and Woodson (1982) hailed item response theory as a cornerstone of progress for educational assessment. Assuming that one can manage the challenges of control and consistency that arise in any study that extends over time, IRT does indeed make it possible to solve many practical problems in assessment, such as allowing item pools to evolve over time, providing results on a consistent scale in the face of complex item-sampling designs, and reducing the numbers of items students are presented.

Possible, but not necessarily easy. Familiar IRT procedures, based on obtaining point estimates for individual examinees, break down in efficient assessments that solicit relatively few responses from each student. This paper and others on IRT in assessment (e.g., Mislevy and Bock, 1988) make it clear that higher levels of theoretical and computational complexity are required to realize the benefits IRT offers.

This paper argues that Rubin's (1987) multiple imputation procedures provide a suitable theoretical framework for latent variables in sample surveys, and illustrates how the procedures can be applied. This method has the advantage of placing the burden of the problem on the primary analyst, who must create completed datasets. With them, the secondary analysts can carry out their research using standard routines for complete data.

References

Andersen, E.B., and Madsen, M. (1977) Estimating the parameters of a latent population distribution. Psychometrika, 42, 357-374.

Baratz-Snowden, J.C., and Duran, R. (1987). The Educational Progress of Language Minority Students: Findings from the 1983-84 NAEP Reading Survey. Princeton: National Assessment of Educational Progress/Educational Testing Service.

Beaton, A.E. (1987). The NAEP 1983/84 Technical Report (NAEP Report 15-TR-20). Princeton: Educational Testing Service.

Beaton, A.E. (1988). The NAEP 1985/86 Technical Report. Princeton: Educational Testing Service.

Beaton, A.E., and Johnson, E.J. (1987). The average response method (ARM) of scoring. In A.E. Beaton, The NAEP 1983/84 Technical Report (NAEP Report 15-TR-20). Princeton: Educational Testing Service.

Bock, R.D., and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. Psychometrika, 46, 443-459.

Bock, R.D., Mislevy, R.J., and Woodson, C.E.M. (1982). The next stage in educational assessment. Educational Researcher, 11, 4-11, 16.

Cassel, C-M, Sarndal, C-E., and Wretman, J.H. (1977). Foundations of Inference in Survey Sampling. New York: Wiley.

Cochran, W.G. (1977). Sampling Techniques. New York: Wiley.

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

Ford, B.L. (1983). An overview of hot-deck procedures. In W.G. Madow, I. Olkin, and D.B. Rubin (Eds.), Incomplete Data in Sample Surveys, Volume 2, Theory and Bibliographies. New York: Academic Press.

Kelly, T.L. (1947). Fundamentals of Statistics. Cambridge: Harvard University Press.

Kirsch, I.S., and Jungeblut, A. (1986). Literacy: Profile of America's Young Adults (Final Report, No. 16-PL-01). Princeton, NJ: National Assessment of Educational Progress.

Laird, N.M. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. Journal of the American Statistical Association, 73, 805-811.

Little, R.J.A., and Rubin, D.B. (1987). Statistical Analysis with Missing Data. New York: Wiley.

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., and Bock, R.D. (1983). BILOG: Item Analysis and Test Scoring with Binary Logistic models [computer program]. Mooresville, IN: Scientific Software, Inc.

Mislevy, R.J., and Bock, R.D. (1988). A hierarchical item response model for educational assessment. In R.D. Bock (Ed.), _Multilevel Analysis of Educational Data_. New York: Academic Press.

Mislevy, R.J., and Sheehan, K.M. (1987). Marginal estimation procedures. In A.E. Beaton, _The NAEP 1983/84 Technical Report_ (NAEP Report 15-TR-20). Princeton: Educational Testing Service.

Mislevy, R.J., and Sheehan, K.M. (in press). The role of collateral information about examinees in the estimation of item parameters. _Psychometrika_.

National Assessment of Educational Progress (1985). _The Reading Report Card: Progress toward Excellence in our Schools_. Princeton: Educational Testing Service.

Rigdon, S. E., and Tsutakawa, R.K. (1983). Parameter estimation in latent trait models. _Psychometrika_, _48_, 567-574.

Rubin, D.B. (1976). Inference and missing data. _Biometrika_, _63_, 581-592.

Rubin, D.B. (1987). _Multiple imputation for nonresponse in surveys_. New York: Wiley.

Sanathanan, L, and Blumenthal, N. (1978) The logistic model and estimation of latent structure. _Journal of the American Statistical Association_, _73_, 794-798.

Sheehan, K.M. (1985). _M-GROUP: Estimation of Group Effects in_

Multivariate Models [computer program]. Princeton: Educational Testing Service.

Zwick, R. (1987). Assessing the dimensionality of NAEP reading data. Journal of Educational Measurement, 24, 293-308.

Table 1

Expectations of Secondary Analyses

	Dependent Variabl				
Population Attribute	θ and $\tilde{\theta}$	$\hat{\theta}$	$\bar{\theta}_{xy}$	$\bar{\theta}_{x}$	$\bar{\theta}_{x}^{(r)}$
Mean	0	0	0	0	0
Variance	1	$1+\sigma_e^2$	$1-\sigma_{\theta\|xy}^2$	$1-\sigma_{\theta\|x}^2 \equiv \rho$	1
Simple Regression					
Coefficient	β_1	β_1	β_1	$\rho \quad \beta_1$	$\sqrt{\rho} \quad \beta_1$
Residual Variance	$\sigma_{\theta\|y1}^2$	$\sigma_{\theta\|y1}^2+\sigma_e^2$	$\sigma_{\theta\|y1}^2-\sigma_{\theta\|xy}^2$	$\rho^2(\sigma_{\theta\|y1}^2+\sigma_e^2)$	$\rho\ (\sigma_{\theta\|y1}^2+\sigma_e^2)$
Proportion of variance accounted for	$1-\sigma_{\theta\|y1}^2$	$\dfrac{1-\sigma_{\theta\|y1}^2}{1+\sigma_e^2}$	$\dfrac{1-\sigma_{\theta\|y1}^2}{1-\sigma_{\theta\|xy}^2}$	$1-\rho(\sigma_{\theta\|y1}^2+\sigma_e^2)$	$1-\rho(\sigma_{\theta\|y1}^2+\sigma_e^2)$

Table 2

Numerical Values for a Short Assessment Instrument

	Dependent Variable				
Population Attribute	θ and $\tilde{\theta}$	$\hat{\theta}$	$\bar{\theta}_{xy}$	$\bar{\theta}_{x}$	$\bar{\theta}_{x}^{(r)}$
Mean	.000	.000	.000	.000	.000
Variance	1.000	2.000	.600	.500	1.000
Simple Regression					
Coefficient	.500	.500	.500	.250	.354
Residual Variance	.750	1.750	.350	.438	.875
% variance accounted for	.250	.125	.417	.125	.125

Table 3

Numerical Values for a Long Test

	Dependent Variable				
Population Attribute	θ and $\tilde{\theta}$	$\hat{\theta}$	$\bar{\theta}_{xy}$	$\bar{\theta}_{x}$	$\bar{\theta}_{x}^{(r)}$
Mean	.000	.000	.000	.000	.000
Variance	1.000	1.100	.940	.910	1.000
Simple Regression					
Coefficient	.500	.500	.500	.455	.477
Residual Variance	.750	.850	.690	.703	.773
% variance accounted for	.250	.227	.266	.227	.227

Table 4

Estimated Proportions Above Cut Point

Population Attribute	Dependent Variable				
	θ and $\tilde{\theta}$	$\hat{\theta}$	$\bar{\theta}_{xy}$	$\bar{\theta}_{x}$	$\bar{\theta}_{x}^{(r)}$
Short Test					
$P(\theta \geq 1)$	.1587	.2611	.0985	.0778	.1587
$P(\theta \geq 1 \mid y_1 = -1)$	.0418	.1292	.0055	.0294	.0735
Long Test					
$P(\theta \geq 1)$	.1587	.1711	.1515	.1469	.1587
$P(\theta \geq 1 \mid y_1 = -1)$	.0418	.0516	.0351	.0409	.0465

Table 5

Expected Regression Coefficients for a Short and a Long Test, with Complete and Incomplete[1] Conditioning for Imputations

Population Attribute	Dependent Variable		
	θ and $\tilde{\theta}$	$\tilde{\theta}^*$ $(\rho_T=.50)$	$\tilde{\theta}^*$ $(\rho_T=.91)$
Simple regression			
β_1	.500	.500	.500
β_2	.500	.357	.471
Multiple regression			
$\beta_{1\mid 2}$	.333	.429	.353
$\beta_{2\mid 1}$	.333	.143	.294

[1] Imputations constructed by conditioning on y_1 but not y_2.

Table 6

Estimates of Conditional Distribution Parameters

Effect	Level	Grade 4/ Age 9	Grade 8/ Age 13	Grade 11/ Age 17
Intercept	All subjects	-1.351	-.433	.159
Gender	Male*	.000	.000	.000
	Female	.096	.139	.160
Ethnicity	Black*	.000	.000	.000
	White & other	.460	.403	.405
	Hispanic	.076	.113	.135
STOC	Rural or Low metro*	.000	.000	.000
	High metro	.490	.308	.230
	Other	.245	.122	.148
Region	Northeast*	.000	.000	.000
	Central	-.133	-.042	.028
	Southeast	-.009	-.021	.023
	West	-.087	-.042	.005
Parent Ed.	Less than HS*	.000	.000	.000
	High school grad	.209	.140	.082
	Beyond HS	.395	.404	.379
	Don't know/missing	.120	-.017	-.075
Grade/Age**	< M age, = M grade*	.000	.000	.000
	= M age, < M grade	-.672	-.433	-.617
	= M age, = M grade	-.065	-.013	-.084
	= M age, > M grade	.338	.549	.077
	> M age, = M grade	-.307	-.260	-.533
Misc.	Subjects with unrecoverable missing values	.510	-.329	.810
Residual variance		.464	.386	.457
Sample size		22,950	23,553	23,932

* Effect fixed at zero

** "M" denotes "modal"; e.g., "> M age, = M grade" means "above the modal age and at the modal grade in one's age/grade cohort."

Table 7

Estimating Age 17 Means from Completed Datasets

	Total	Males	Black	West	Rural
(1) Imputation 1	288.005	282.644	266.195	287.177	282.990
(2) Imputation 2	288.258	283.201	265.104	288.338	283.499
(3) Imputation 3	288.208	282.869	265.259	288.018	283.285
(4) Imputation 4	288.135	282.554	264.832	287.745	282.854
(5) Imputation 5	287.819	282.314	264.241	287.196	282.360
(6) Average (1)-(5)	288.085	282.718	265.126	287.695	282.997
(7) Variance (1)-(5)	.025	.092	.406	.208	.152
(8) Average jackknife variance	1.248	1.225	1.742	4.333	9.218
(9) Total variance 1.2 x (7) + (8)	1.278	1.335	2.229	4.583	9.400
(10) Proportional increase [(9)-(8)]/(8)	.024	.090	.280	.058	.020

Table 8

Multiple Regression Estimates based on Imputations
Constructed with Partial and Full Conditioning

	Partial	Full	Conditioning		%-attenuation	
Effect	β	β	SE(β)	t	all	significant
White; language minority	6.08	4.23	3.96	1.07	-43.74	
White; language non-minority	12.22	13.72	2.94	4.67	10.93	10.93
Hispanic; lang. non-minority	-.76	1.22	3.25	.38	162.30	
Asian; language minority	-2.90	-6.25	4.39	-1.42	53.60	
Asian; language non-minority	9.54	17.34	4.66	3.72	44.98	44.98
Black; language non-minority	-8.64	-10.82	2.95	-3.67	20.15	20.15
Sex = male*	-8.55	-9.35	.80	-11.69	8.56	8.56
Parent education*	6.03	5.80	.38	15.26	-3.97	-3.97
Home language minority	-9.41	-13.78	2.78	-4.96	31.71	31.71
Study aids	2.63	3.89	.43	9.05	32.39	32.39
Homework	2.78	3.82	.30	12.73	27.23	27.23
Hours of TV	-1.22	-2.04	.24	-8.50	40.20	40.20
Pages read	6.36	10.59	1.01	10.49	39.94	39.94
Years academic courses	.91	1.25	.14	8.93	27.20	27.20

* Effect included in partial conditioning set

Dr. Terry Ackerman
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

Dr. Robert Ahlers
Code N711
Human Factors Laboratory
Naval Training Systems Center
Orlando, FL 32813

Dr. James Algina
1403 Norman Hall
University of Florida
Gainesville, FL 32605

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Isaac Bejar
Mail Stop: 10-R
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Ramat Aviv 69978
ISRAEL

Dr. Arthur S. Blaiwes
Code N712
Naval Training Systems Center
Orlando, FL 32813-7100

Dr. Bruce Bloxom
Defense Manpower Data Center
550 Camino El Estero,
Suite 200
Monterey, CA 93943-3231

Dr. R. Darrell Bock
University of Chicago
NORC
6030 South Ellis
Chicago, IL 60637

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code 7B
Naval Training Systems Center
Orlando, FL 32813-7100

Dr Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Dr. James Carlson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd., North
Chapel Hill, NC 27514

Dr. Robert M. Carroll
Chief of Naval Operations
OP-01B2
Washington, DC 20350

Dr. Raymond E. Christal
UES LAMP Science Advisor
AFHRL/MOEL
Brooks AFB, TX 78235

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
Los Angeles, CA 90089-1061

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Hans F. Crombag
Faculty of Law
University of Limburg
P.O. Box 616
Maastricht
The NETHERLANDS 6200 MD

Dr. Timothy Davey
Educational Testing Service
Princeton, NJ 08541

Dr. C. M. Dayton
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Ralph J. DeAyala
Measurement, Statistics,
and Evaluation
Benjamin Bldg., Rm. 4112
University of Maryland
College Park, MD 20742

Dr. Dattprasad Divgi
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong
Bell Communications Research
6 Corporate Place
PYA-1K226
Piscataway, NJ 08854

Dr. Fritz Drasgow
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
(12 Copies)

Dr. Stephen Dunbar
224B Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. John M. Eddins
University of Illinois
252 Engineering Research
Laboratory
103 South Mathews Street
Urbana, IL 61801

Dr. Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence, KS 66045

Dr. George Englehard, Jr.
Division of Educational Studies
Emory University
210 Fishburne Bldg.
Atlanta, GA 30322

Dr. Benjamin A. Fairbank
Performance Metrics, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

Dr. P-A. Federico
Code 51
NPRDC
San Diego, CA 92152-6800

Dr. Leonard Feldt
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
P.O. Box 168
Iowa City, IA 52243

Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Dr. Myron Fischl
U.S. Army Headquarters
DAPE-MRR
The Pentagon
Washington, DC 20310-0300

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Alfred R. Fregly
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Robert D. Gibbons
Illinois State Psychiatric Inst.
Rm 529W
1601 W. Taylor Street
Chicago, IL 60612

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

DORNIER GMBH
P.O. Box 1420
D-7990 Friedrichshafen 1
WEST GERMANY

Dr. Ronald K. Hambleton
University of Massachusetts
Laboratory of Psychometric
and Evaluative Research
Hills South, Room 152
Amherst, MA 01003

Dr. Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Dr. Grant Henning
Senior Research Scientist
Division of Measurement
Research and Services
Educational Testing Service
Princeton, NJ 08541

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 63
San Diego, CA 92152-6800

Dr. Paul W. Holland
Educational Testing Service, 21-T
Rosedale Road
Princeton, NJ 08541

Prof. Lutz F. Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 92010

Mr. Dick Hoshaw
OP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
3-104 Educ. N.
University of Alberta
Edmonton, Alberta
CANADA T6G 2G5

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Robert Jannarone
Elec. and Computer Eng. Dept.
University of South Carolina
Columbia, SC 29208

Dr. Douglas H. Jones
Thatcher Jones Associates
P.O. Box 6640
10 Trafalgar Court
Lawrenceville, NJ 08648

Dr. Milton S. Katz
European Science Coordination
Office
U.S. Army Research Institute
Box 65
FPO New York 09510-1500

Prof. John A. Keats
Department of Psychology
University of Newcastle
N.S.W. 2308
AUSTRALIA

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
Box 7246, Meas. and Eval. Ctr.
University of Texas-Austin
Austin, TX 78703

Dr. James Kraatz
Computer-based Education
Research Laboratory
University of Illinois
Urbana, IL 61801

Dr. Leonard Kroeker
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Dr. Jerry Lehnus
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn, VA 22209

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, WI 53705

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541-0001

Dr. Robert L. Linn
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Dr. Robert Lockman
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. George B. Macready
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. James R. McBride
The Psychological Corporation
1250 Sixth Avenue
San Diego, CA 92101

Dr. Clarence C. McCormick
HQ, USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago, IL 60064

Dr. Robert McKinley
Educational Testing Service
16-T
Princeton, NJ 08541

Dr. James McMichael
Technical Director
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Barbara Means
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. William Montague
NPRDC Code 13
San Diego, CA 92152-6800

Ms. Kathleen Moreno
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Headquarters Marine Corps
Code MPI-20
Washington, DC 20380

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Norman, OK 73071

Deputy Technical Director
NPRDC Code 01A
San Diego, CA 92152-6800

Director, Training Laboratory,
NPRDC (Code 05)
San Diego, CA 92152-6800

Director, Manpower and Personnel
Laboratory,
NPRDC (Code 06)
San Diego, CA 92152-6800

Director, Human Factors
& Organizational Systems Lab,
NPRDC (Code 07)
San Diego, CA 92152-6800

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

Commanding Officer,
Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 801
Department of Educational
Psychology & Technology
University of Southern California
Los Angeles, CA 90089-0031

Dr. James B. Olsen
WICAT Systems
1875 South State Street
Orem, UT 84058

Office of Naval Research,
Code 1142CS
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Off.ce of Naval Research,
Code 125
800 N. Quincy Street
Arlington, VA 22217-5000

Assistant for MPT Research,
Development and Studies
OP 01B7
Washington, DC 20370

Dr. Judith Orasanu
Basic Research Office
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. Randolph Park
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dept. of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey, CA 93943-5026

Department of Operations Research,
Naval Postgraduate School
Monterey, CA 93940

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MOA
Brooks AFB, TX 78235

Dr. Barry Riegelhaupt
HumRRO
1100 South Washington Street
Alexandria, VA 22314

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg.
Knoxville, TN 37916-0900

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Mary Schratz
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Dan Segall
Navy Personnel R&D Center
San Diego, CA 92152

Dr. W. Steve Sellman
OASD(MRA&L)
2B269 The Pentagon
Washington, DC 20301

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr. William Sims
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. H. Wallace Sinaiko
Manpower Research
and Advisory Services
Smithsonian Institution
801 North Pitt Street, Suite 120
Alexandria, VA 22314-1713

Dr. Richard E. Snow
School of Education
Stanford University
Stanford, CA 94305

Dr. Richard C. Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Paul Speckman
University of Missouri
Department of Statistics
Columbia, MO 65201

Dr. Judy Spray
ACT
P.O. Box 168
Iowa City, IA 52243

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
Code-62
San Diego, CA 92152-6800

Dr. John Tangney
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
103 S. Mathews Avenue
Urbana, IL 61801

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Robert Tsutakawa
University of Missouri
Department of Statistics
222 Math. Sciences Bldg.
Columbia, MO 65211

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E. Street, NW
Washington, DC 20415

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 440
St. Paul, MN 55114

Dr. Frank L. Vicino
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

Dr. Ming-Mei Wang
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Thomas A. Warm
Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Dr. Brian Waters
HumRRO
12908 Argyle Circle
Alexandria, VA 22314

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Ronald A. Weitzman
Box 146
Carmel, CA 93921

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Douglas Wetzel
Code 51
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90089-1061

German Military Representative
ATTN: Wolfgang Wildgrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

Dr. Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana, IL 61801

Dr. Hilda Wing
NRC MH-176
2101 Constitution Ave.
Washington, DC 20418

Dr. Martin F. Wiskoff
Defense Manpower Data Center
550 Camino El Estero
Suite 200
Monterey, CA 93943-3231

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering
Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
Code 51
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
03-T
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

Dr. Joseph L. Young
National Science Foundation
Room 320
1800 G Street, N.W.
Washington, DC 20550

Mr. Anthony R. Zara
National Council of State
  Boards of Nursing, Inc.
625 North Michigan Avenue
Suite 1544
Chicago, IL 60611

Dr. Peter Stoloff
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268