DOCUMENT RESUME

ED 348 399 TM 018 884

AUTHOR	Junker, Brian W.
TITLE	A Note on Recovering the Ability Distribution from Test Scores.
INSTITUTION	Carnegie Mellon Univ., Pittsburgh, PA. Dept. of Statistics.
SPONS AGENCY	Office of Naval Research, Arlington, VA. Cognitive and Neural Sciences Div.
REPORT NO	ONR/CS-92-1
PUB DATE	May 92
CONTRACT	N00014-91-J-1208
NOTE	36p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Computer Simulation; *Equations (Mathematics); *Estimation (Mathematics); Item Bias; Item Response Theory; *Mathematical Models; *Scores; *Statistical Distributions; Test Length
IDENTIFIERS	*Ability Estimates; Local Independence (Tests); Population Parameters; Smoothing Methods; *Theta Estimates

ABSTRACT

A simple scheme is proposed for smoothly approximating the ability distribution for relatively long tests, assuming that the item characteristic curves (ICCs) are known or well estimated. The scheme works for a general class of ICCs and is guaranteed to completely recover the theta distribution as the test length increases. The proposed method of estimating the ability distribution is robust to some violations of local independence. After an initial function inversion, the scheme can be inexpensively used to recover the theta distribution in each of several different administrations of the same test or several subpopulations in one test administration. Moreover, this approach could be used to recover the distribution of a dominant ability dimension when local independence fails. The scheme provides a starting place for diagnostics concerning assumptions about the shape of the theta distribution or ICCs of a particular test. Work is currently under way to further examine and refine these methods using essentially unidimensional simulation data and to apply the estimator to real tests. Kernel smoothing is also considered. A 16-item list of references, 10 tables, 8 graphs, and 2 appendixes that provide details of the simulation and proofs are included. (RLC)

A Note on Recovering the Ability Distribution from Test Scores

by

Brian W. Junker

A Note on Recovering the Ability Distribution from Test Scores

by

Brian W. Junker

Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213

May 1992

Technical Report ONR/CS 92-1

Prepared for the Model-Based Measurement Program, Cognitive and Neural Sciences Division, Office of Naval Research, under grant number N00014-91-J-1208, R&T 4421-560. Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	20 May 1992	Technical

4. TITLE AND SUBTITLE

A Note on Recovering the Ability Distribution from Test Scores.

5. FUNDING NUMBERS

G: N00014-91-J-1208
PE: 61153N
PR: RR04204
TA: RR04204-01
WU: 4421-560-4

6. AUTHOR(S)

Brian W. Junker

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213

8. PERFORMING ORGANIZATION REPORT NUMBER

ONR/CS 92-1

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

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

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Submitted for publication

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We propose a simple scheme for smoothly approximating the ability distribution for relatively long tests, assuming that the ICC's are known or well estimated. The scheme works for quite a general class of item characteristic curves (ICC's) and is guaranteed to completely recover the Θ distribution as the test length, J, grows. After an initial function inversion, the scheme can be inexpensively used to recover the Θ distribution in each of several different administrations of the same test (or subpopulations in one test administration). Moreover, this approach could be used to recover the distribution of a dominant ability dimension when local independence fails. Finally, the scheme provides a starting place for diagnostics concerning assumptions about the shape of the Θ distribution or ICC's of a particular test. Work is currently underway to further examine and refine these methods using essentially unidimensional simulation data, and to apply the estimators to real tests.

14. SUBJECT TERMS

Item response theory, kernel smoothing, latent trait distribution, population assessment.

15. NUMBER OF PAGES

27

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Abstract

We propose a simple scheme for smoothly approximating the ability distribution for relatively long tests, assuming that the ICC's are known or well estimated. The scheme works for quite a general class of item characteristic curves (ICC's) and is guaranteed to completely recover the Θ distribution as the test length, J, grows. After an initial function inversion, the scheme can be inexpensively used to recover the Θ distribution in each of several different administrations of the same test (or subpopulations in one test administration). Moreover, this approach could be used to recover the distribution of a dominant ability dimension when local independence fails. Finally, the scheme provides a starting place for diagnostics concerning assumptions about the shape of the Θ distribution or ICC's of a particular test. Work is currently underway to further examine and refine these methods using essentially unidimensional simulation data, and to apply the estimators to real tests.

Keywords: Item response theory, kernel smoothing, latent trait distribution, population assessment.

The work reported here was initiated under the direction of Paul Holland, while Junker was a participant in the Educational Testing Service Summer Predoctoral Research Program. Initial computer simulations were performed by Dorothy Thayer at ETS; the simulations reported here were performed by Junker at the University of Illinois and Carnegie Mellon University.

1 The basic estimator

A principal application of educational testing is inferring the distribution of abilities in various populations. This task is important for both users of these tests (in, say, comparing various subpopulations) and researchers and test developers (in, say, developing or using item calibration—ICC parameter estimation—procedures within the IRT framework).

Inference about the ability distribution from item response data goes back at least to Lord (1953) who gives an interesting qualitative account of the possible distortions induced by the traditional IRT model. With the rise in popularity of item response theory, IRT, many methods for estimating the latent distribution have been developed.

Samejima and Livingston (1979) fit polynomials to latent densities using the method of moments. Samejima (1984) also fits Θ densities, given the MLE $\hat{\theta}$, using specific parametric families by matching two or more moments. Levine (1984, 1985) projects the (unknown) latent distribution onto a convenient function space in the span of the test's conditional likelihood functions and estimates the projection by maximum likelihood. Mislevy (1984) assumes that the ability distribution is well approximated by a collection of masses centered at points placed a priori along the θ axis and estimates the sizes of the masses at each point. More generally, hierarchical and/or empirical Bayes techniques may be used to estimate parameters of the latent trait distribution if it belongs to a tractable family of priors. These methods all rely upon local independence for their validity; moreover they tend to be expensive in terms of computation and storage.

We will examine a simpler method of estimating the ability distribution which, in addition, is robust to some violations of local independence. Consider a set of J binary items

$$\underline{X}_J \equiv (X_1, X_2, \ldots, X_J)$$

that may be embedded in a longer sequence or pool of items $(X_1, X_2, X_3, \ldots)$. Let Θ be the latent trait of interest, let $P_1(\theta), P_2(\theta), \ldots, P_J(\theta)$ be the item characteristic curves, ICC's, with respect to Θ, and denote averages of items as $\overline{X}_J = \frac{1}{J}\sum_1^J X_j$, and similarly for averages $\overline{P}_J(\theta)$ of ICC's. Under the usual local independence (LI) and monotonicity (M) conditions of item response theory (e.g. Hambleton, 1989), or more generally under Stout's (1990) formulation of essential independence (EI) and local asymptotic discrimination (LAD), we know that $\tilde{\theta}_J(\underline{X}_J) \equiv \overline{P}_J^{-1}(\overline{X}_J)$ is a plausible point estimate of Θ: $\tilde{\theta}_J(\underline{X}_J)$ is a consistent estimator of Θ under either set of assumptions. It immediately follows that the distribution of $\tilde{\theta}_J(\underline{X}_J)$

$$F_J(t) = P[\tilde{\theta}_J(\underline{X}_J) \leq t]$$

converges to that of Θ as well (e.g. Serfling, 1980, p. 19). Now consider administering the test $\underline{X}_J$ to N examinees, obtaining N response vectors $\underline{X}_{1J}, \ldots, \underline{X}_{NJ}$ and corresponding θ estimates $\tilde{\theta}_J(\underline{X}_{1J}), \ldots, \tilde{\theta}_J(\underline{X}_{NJ})$; a natural estimator of the Θ distribution is the "empirical" distribution of these $\tilde{\theta}_J$'s

$$\begin{aligned} \tilde{F}_{N,J}(t) &\equiv \frac{1}{N}\sum_{n=1}^{N} 1_{\{\tilde{\theta}_J(\underline{X}_{nJ})\leq t\}} \qquad (1) \\ &= \left\{\text{fraction of } \tilde{\theta}_J(\underline{X}_{nJ})\text{'s} \leq t\right\} \end{aligned}$$

where the "indicator function" 1_S takes the value 1 if S is true and 0 if S is false.

Theorem 1 *Suppose* $(X_1, X_2, \ldots)$ *is a sequence of items and* Θ *is a latent trait such that EI and LAD hold. Define* $\tilde{\theta}_J(\underline{X}_J)$ *as above. If the distribution function*

$$F(t) = P[\Theta \leq t]$$

is continuous, the empirical distribution function $\tilde{F}_{N,J}(t)$ *defined in (1), converges in probability to* F *at each* t *as both* $J \to \infty$ *and* $N \to \infty$.

As with the work of Stout (1990) and Junker (1991), the embedding in an infinite-length item pool is partly a conceptual tool. In practice, one might check the EI condition using Stout's (1987) test, and check the LAD condition by verifying that the average ICC for a particular test was an invertible function.

In fact, the full strength of the LAD condition is not needed here. A weaker condition that also gives the theorem is that, for all $t_2 > t_1$ there exists $\epsilon(t_1, t_2)$ such that

$$\liminf_{J\to\infty} \overline{P}_J(t_2) - \overline{P}_J(t_1) \geq \epsilon(t_1, t_2)\cdot \tag{2}$$

Similarly, the full strength of the EI condition is not needed. It suffices to have, for all t,

$$\lim_{J\to\infty} \mathrm{Var}(\overline{X}_J|\Theta = t) = 0. \tag{3}$$

Under the weaker conditions (2) and (3), the consistency of $\overline{P_J}^{-1}(\overline{X}_J)$ as a point estimate for θ may fail, but Theorem 1 still goes through. The proof of Theorem 1 is based on a well-known exponential bound due to Dvoretsky, Kiefer and Wolfowitz (Serfling, 1980, p. 59) on the error made in approximating $F_J(t)$ with $\tilde{F}_{N,J}(t)$. See Appendix B for some details.

2 Two practical considerations

Note that the theorem does not in any way require that the ICC's have 0 and 1 as lower and upper asymptotes. For example, if $\overline{P}_J$ has a lower asymptote c, i.e.,

$$\liminf_{J\to\infty} \overline{P}_J(t) > c \geq 0, \forall t \in \mathbb{R},$$

there certainly could be positive probability that some $\underline{X}_J$'s have $\overline{X}_J \leq c$. The only reasonable thing for $\overline{P_J}^{-1}$ to do with such an $\overline{X}_J$ is send it to $-\infty$, which ruins the estimate of F.

But for any fixed θ, if $c < \liminf_{J\to\infty} \overline{P}_J(\theta)$,

$$\begin{aligned}
\limsup_{J\to\infty} P[\overline{X}_J \leq c] &= \limsup_{J\to\infty} \int_{-\infty}^{\infty} P[\overline{X}_J \leq c|\Theta = t]dF(t) \\
&\leq \limsup_{J\to\infty} \int_{-\infty}^{\infty} P[\overline{X}_J \leq \overline{P}_J(\theta)|\Theta = t]dF(t) \\
&= F(\theta),
\end{aligned}$$

after observing that $P[\overline{X}_J \leq \overline{P}_J(\theta)|\Theta = t] \rightarrow 1_{\{t \leq \theta\}}$ and applying standard convergence results (Ash, 1972). By letting $\theta \rightarrow -\infty$ it follows that

$$\lim_{J \to \infty} P[\overline{X}_J \leq c] = 0.$$

The distribution of $\tilde{\theta}_J(\underline{X}_J)$ does indeed place mass at $-\infty$ for some scores (e.g., for $\overline{X}_J/J = 0$ and fails to "recover" the Θ distribution for those scores. The point of the calculation is that as J grows, the part of the Θ distribution corresponding to these "bad" scores becomes negligible, so we don't have to worry, theoretically, about its not being recovered. Indeed, under local independence, we can further calculate that $P[\underline{X}_J \leq c]$ falls off essentially geometrically as $J \rightarrow \infty$ (Hoeffding 1963, p. 15).

However in practice we still must be concerned about $\overline{X}_J$'s below a lower asymptote c, or above an upper asymptote d. In the pilot simulation described below we have made two adjustments for this problem. Our first adjustment replaces the basic point estimate $\tilde{\theta}_J$ with an estimator based on a shrunken $\overline{X}_J$:

$$\tilde{\theta}_J^{(1)}(\underline{X}_J) = \overline{P}_J^{-1}\left[\frac{J \cdot \overline{X}_J + 1}{J + 2}\right].$$

This estimator also converges in distribution to Θ , and it is evidently bounded (for fixed J) if the asymptotes of $\overline{P}_J$ are 0 and 1. Our second adjustment is in the numerical inversion of the function $\overline{P}_J$ on the computer. We have written the inverter (a secant variation of Newton's method) so that it finds a root of a linear extrapolation of $\overline{P}_J(t) = \overline{X}_J$ when $\overline{X}_J$ lies outside the asymptotes of $\overline{P}_J$. This adjustment is innocuous asymptotically.

Finally, note that this method (like others) requires "perfect" knowledge of the ICC's. In practice of course one never knows the ICC's perfectly, so it is important to know what happens if the "wrong" ICC's are used in the definition of $\tilde{\theta}_J$. For example, how sensitive is this method to using estimates of the item parameters in a 3PL (three parameter logistic ICC) model. instead of the true parameters; or how far off is the estimated Θ distribution if the true ICC's are 3PL's. but only Rasch ICC's are used to calculate $\tilde{\theta}_J$?

Theorem 2 *Suppose $X_1, X_2, \ldots$ and Θ are as in Theorem 1 with ICC's $P_1(t)$, $P_2(t)$, ..., with average $\overline{P}_J(t)$ as before, and suppose*

$$R_1(t),\ R_2(t),\ \ldots$$

are another set of ICC's, with average $\overline{R}_J(t)$. Let $\overline{P_J}^{-1}$ and $\overline{R_J}^{-1}$ be the corresponding inverses, and let

$$\tilde{\theta}_J(\underline{X}) = \overline{R}_J^{-1}(\overline{X}_J).$$

Fix θ such that $\overline{P_J}^{-1}\overline{R}_J(\theta)$ has a finite limit $\tau(\theta)$. Then

$$F_J(\theta) = P[\tilde{\theta}_J(\underline{X}_J) \le \theta] \to F(\tau(\theta))$$

(where F is the distribution of Θ). If these hypotheses hold for every θ, and if τ and F are continuous functions, then the convergence is uniform in θ.

The existence of the limit $\tau(\theta)$ is a technical requirement that, like LAD, is innocuous in the context of real, finite length tests. The most useful interpretation of Theorem 2 is that

$$|F_J(\theta) - F[\overline{P_J}^{-1}\overline{R}_J(\theta)]| \to 0$$

as $J \to \infty$, i.e., the distribution of Θ is estimated with a distortion $\overline{P_J}^{-1}\overline{R}_J$. This follows from the theorem if F is continuous at $\tau(\theta)$.

The proof of Theorem 2 expands on the technique used to prove convergence of $F_J(\theta)$ to $F(\theta)$; see Appendix B. Just as in Theorem 1 it is also possible to show that the empirical distributions

$$\tilde{F}_{N,J}(t) = \frac{1}{N}\sum_{n=1}^{N} 1_{\{\tilde{\theta}_J(\underline{X}_{J,1}) \le t\}}$$

converge to $F(\tau(\theta))$.

The value of Theorem 2 is that if the function $\overline{P}_J^{-1}(\overline{R}_J(\theta))$ can be (partially) identified, then the distribution of $\tilde{\theta}_J$ can still tell us a lot about the underlying Θ distribution. For example, if the "true ICC's" are $P_j(\theta)$ and the Θ distribution is recovered with "estimated ICC's" $R_j(\theta)$, with the estimated ICC's satisfying

$$|\overline{P}_J(\theta) - \overline{R}_J(\theta)| \to 0$$

as $J \to \infty$, then the estimated distributions F_J will converge to the true distribution F of Θ, as long as the derivative $\overline{P}'_J(\theta)$ is bounded away from zero at each θ as $J \to \infty$ (this is guaranteed by LAD for example).

Some knowledge of the underlying Θ distribution may even be available when the "true ICC's" $P_j(\theta)$ and the "recovery ICC's" $R_j(\theta)$ do not match up asymptotically. For example, it is easy to check numerically that for "typical" parameter values, averages of logistic ICC's are themselves approximately logistic (with parameters approximately the averages of the discrimination and difficulty parameters of the individual ICC's). Thus for example if the $P_j(\theta)$ are Rasch (one-parameter logistic) and the estimation method for the "difficulty parameters" b_j is known, on average, to bias the $\hat{b}_j$ by some fixed but unknown additive bias parameter β (so that $logit\ R_j(\theta) \approx logit\ P_j(\theta) + \beta$) then roughly $\overline{P}_J^{-1}(\overline{R}_J(\theta)) \approx \alpha\theta - \beta$, with α near 1, so that the location of the Θ distribution will be estimated wrongly but the (shape) family to which it belongs may still be identified. Similar considerations apply when the $P_j(\theta)$ are 3PL, and the $R_j(\theta)$ are 2PL: over the domain of $\overline{P}_J^{-1}(\theta)$, $\overline{P}_J^{-1}(\overline{R}_J(\theta))$ is approximately linear.

3 Kernel smoothing

The basic estimator proposed in (1) is the "empirical distribution" function

$$\begin{aligned} \hat{F}_{N,J}(t) &= \frac{1}{N}\sum_{n=1}^{N} 1_{\{\overline{P}_J^{-1}(\overline{X}_{nJ}) \le t\}} \\ &= \sum_{j=0}^{J} \hat{P}_N[\overline{X}_J = j/J]\ 1_{\{\overline{P}_J^{-1}(j/J) \le t\}} \end{aligned} \tag{4}$$

where

$$\hat{P}_N\{\overline{X}_J = j/J\} = \frac{1}{N}\sum_{n=1}^{N} 1_{\{\overline{X}_{nJ}=j/J\}}$$

is the natural estimator of the (discrete) distribution of $\overline{X}_J$ based on N observations $\overline{X}_{1J}$, ..., $\overline{X}_{NJ}$. The indicator function on the far right in (4) may be written

$$1_{\{\overline{P}_J^{-1}(j/J)\leq t\}} = \tilde{K}\left[\frac{t-\overline{P}_J^{-1}(j/J)}{h}\right],$$

where $\tilde{K}(u)$ is constant, except for a jump from 0 to 1 at $u = 0$, and h is any positive number. In cases where the Θ distribution F is continuous, we may be able to improve the performance of $\tilde{F}_{N,J}$ by replacing the discrete function $\tilde{K}$ with a continuous distribution function $K(u)$ increasing from 0 to 1 as u ranges from $-\infty$ to ∞. Denote the smoothed estimator as

$$\begin{aligned}\hat{F}_{NJh}(t) &= \sum_{j=0}^{J}\hat{P}_N[\overline{X}_J = j/J]K\left[\frac{t-\overline{P}_J^{-1}(j/J)}{h}\right]\\ &= \frac{1}{N}\sum_{n=1}^{N}K\left[\frac{t-\overline{P}_J^{-1}(\overline{X}_{nJ})}{h}\right].\end{aligned} \tag{5}$$

This estimator is in the same spirit as kernel density estimators for estimating the density of a continuous random variable V based on direct, independent observations $V_1, V_2, \ldots, V_N$:

$$\hat{f}_N(t) = \frac{1}{nh}\sum_{n=1}^{N}k\left[\frac{t-V_n}{h}\right]$$

where $k(t)$ is a fixed density (see for example Silverman, 1986). However it differs from these estimators in several ways.

First, our estimator $\hat{F}_{NJh}$ is a distribution estimator, not a density estimator. Reiss (1981) is another example in which kernel smoothing is used to estimate distributions.

Second, we are not allowed direct access to the observations $\Theta_1, \ldots, \Theta_N$. We must base our estimation of F on the discrete, noisy transformations $\overline{X}_{1J}, \ldots, \overline{X}_{NJ}$ of $\Theta_1, \ldots, \Theta_N$. Note that the "granularity" of these observations changes with J.

Third, the observations $\overline{X}_{1J}, \ldots, \overline{X}_{NJ}$ must be transformed by the nonlinear transformation $\overline{P}_J^{-1}$. This means that the granularity changes over the range of Θ and $\overline{X}_J$; this complicates practical calculations such as those leading to optimal rates for N, J and h.

We now show that the weighted root mean square error (RMS) between this estimator and the true Θ distribution goes to zero as $N, J \to \infty$. The theorem below is analogous to Theorem 1.

Theorem 3 *Suppose $X_1, X_2, \ldots$ and Θ are as in Theorem 1 with ICC's $P_1(\theta), P_2(\theta), \ldots$ Define $\hat{F}_{NJh}(t)$ as in (5), for a fixed kernel distribution function K. Then if the distribution function F of Θ is continuous, and K has a finite first absolute moment,*

$$RMS \equiv \left\{ E \int_{-\infty}^{\infty} [\hat{F}_{NJh}(t) - F(t)]^2 g(t)dt \right\}^{1/2} \to 0 \tag{6}$$

as $N \to \infty, J \to \infty$ and $h \to 0$, for any density $g(t)$.

Unlike most nonparametric density estimation results, there is no restriction on the rates at which $h \to 0$, $N \to \infty$ or $J \to \infty$. This is partly because a distribution function is smoother than, and therefore easier to estimate than, a density. The corresponding technique for estimation of the Θ *density* would require h^3 to tend to zero more slowly than $E[\tilde{\theta}_J(\overline{X}_J) - \Theta]$, for example, as well as further conditions on the rates at which N and J tend to ∞. Despite the fact that there are no rates in the theorem, devising h as a function of N and J to produce the "right" amount of smoothing is an important issue to which we shall return below.

The proof of Theorem 3 (see Appendix B) is based on decomposing the RMS in (6) as

$$\begin{aligned} RMS^2 &= \int_{-\infty}^{\infty} \{P[\overline{P}_J^{-1}(\overline{X}_J) + hY \le t] - P[\Theta \le t]\}^2 g(t)dt \\ &\quad + \frac{1}{N} \int_{-\infty}^{\infty} \operatorname{Var} K \left[\frac{t - \overline{P}_J^{-1}(\overline{X}_J)}{h} \right] g(t)dt \end{aligned} \tag{7}$$

where Y is a random variable with distribution K, independent of Θ and all item responses. This technique can be modified to show that

$$E[\hat{F}_{NJh}(t) - F(t)]^2 \to 0$$

for any t, and hence $\hat{F}_{NJh}(t) \to F(t)$ in probability, for each continuity point t of F. For example, this provides another proof that our original estimator $\hat{F}_{N,J}$ converges in probability to F. It would also be clear from the proof that the same smoothing could be applied with any consistent estimator $\tilde{\theta}_J$ in place $\overline{P}_J^{-1}(\overline{X}_J)$.

From the decomposition of RMS in (7) into squared-bias and variance terms it seems that the optimal h should be more sensitive to J than N. Indeed, when J is small and N is relatively large, the coarse granularity inherent in $\overline{P}_J^{-1}(\overline{X}_J)$ should predominate over the finer granularity inherent in observing N examinees.

A workable approach to setting h is to make a quick, crude estimate of the variance of Θ by assuming that $\overline{X}_J$ is uniformly distributed on the interval defined by the lower asymptote c and the upper asymptote d of $\overline{P}_J(\theta)$ and then applying the formula

$$h = C \cdot J^{-1/5} \cdot (\mathrm{Var}\,\Theta)^{1/2} \tag{8}$$

which seems appropriate when K has a variance (Silverman, 1986, pp. 45–48; Reiss, 1981). Our crude estimate of Var Θ is obtained by tabulating values of $\tilde{\theta}_j^{(1)} = \overline{P}_J^{-1}((j+1)/(J+2))$ for all j such that $c < (j+1)/(J+2) < d$, and calculating

$$(\mathrm{Var}\,\Theta)^{1/2} \approx (.7413)(\textit{interquartile range})$$

(following the relationship between interquartile range and standard deviation for the Normal distribution). Preliminary trials with $C = 1, 1/2, 1/3, 1/4$ in (8) indicated that $C = 1/3$ produced the best RMS results.

There is reason to believe that choice of K should not be very influential on the RMS in (6) (Silverman, 1986, pp. 42-43). The K used in our simulations was

$$K(t) \;=\; \int_{-\infty}^{t} \frac{3}{4}(1-u^2)\; 1_{\{|u|<1\}} du$$

$$= \begin{cases} 0 & , \quad t < -1 \\ \frac{1}{4}(3t - t^3 + 2) & , \quad |t| \le 1 \\ 1 & , \quad t > 1 \end{cases} \quad . \tag{9}$$

This choice is conservative about the tails of the Θ distribution.

4 Computer simulation

The estimators proposed in Theorems 1 through 3 are less complicated than distribution estimators currently in use in IRT. To help evaluate these estimators a pilot simulation study was performed. In this simulation, item response data was generated using various $d_L = 1$ parametric models, and we attempted to recover the ability distribution using both the smoothed and unsmoothed estimators.

Monte Carlo trials:	$M = 100$	
Examinee sample size:	$N = 5,000$	
Ability distribution:	Normal	$N(0,1)$
	Bimodal Mixture	$\frac{1}{2}N(-1.5,1) + \frac{1}{2}N(1.5,1)$
	Discontinuous	$\chi^2_1 - 1$
Test length:	$J = 10, 30, 60, 100$	
ICC type:	Rasch:	b_j's equally spaced from −2 to 2
	3PL:	b_j's equally spaced from −2 to 2
		a_j's cycling through 0.5, 1.0, 1.5
		c_j's all set to 0.2
	'Estimated':	Generated with the 3PL ICC's above;
		Estimated with the ICC parameters:
		$\beta_j \sim N(b_j, 1/J)$
		$\alpha_j \sim N(a_j, 0.25)$
		$\gamma_j \sim \max\{N(0.2, 0.1), 0\}$
		(all independent).

Table 1: Monte Carlo simulation parameters.

The parameters of the pilot simulation are indicated in Table 1. All possible combinations of these parameters were investigated. The choice of ability distributions was intended to examine two "typical" and one "worst case" target distribution. While the standard normal distribution is extremely smooth and has a bounded positive density the distribution of the shifted chi-squared random variable $\chi_1^2 - 1$ puts no mass below $\theta = -1$ and the density jumps from 0 to $+\infty$ at $\theta = -1$. (This choice is not intended to be terribly realistic, but allows us to explore the performance of our distribution estimator under adverse circumstances.) Although the means of these distributions are both 0, the chi-squared distribution has twice the variance of the normal. The bimodal mixture was chosen to represent a situation where two radically different types of examinee take the test. Its standard deviation is also greater than 1 (roughly 1.8).

The ICC's used were all subfamilies of the three parameter logistic (3PL) curves:

$$P_j(t) = c_j + (1 - c_j)[1 + \exp[-a_j[t - b_j]]^{-1}.$$

In the case labelled "Rasch", $a_j \equiv 1, c_j \equiv 0$ and b_j are as indicated. The same ICC's were used to recover F as to generate the data. Indeed $\tilde{\theta}_J^{(1)}$ is exactly the MLE for θ under the Rasch model with known item parameters. Similarly for the 3PL case, where all the parameters were allowed to vary as indicated above; now $\tilde{\theta}_J^{(1)}$ is a somewhat inefficient estimator of θ. In the case labelled 'Estimated', the 3PL ICC's were used to generate the data ($P_j(\theta)$'s in Theorem 2) but then their item parameters were deliberately contaminated with noise to produce the "recovery ICC's" ($R_j(\theta)$'s in Theorem 2) used to estimate F, to roughly approximate the practical situation in which item parameters themselves must be estimated from data. Thus the cases Rasch, 3PL, and 'Estimated' represent increasingly hostile situations for the distribution estimator to work in.

Finally, the choice of N = 5,000 examinees was somewhat arbitrary. In preliminary runs, N = 1,000 and N = 10,000 yielded measures of fit of the estimated ability distribution to the true distribution quite comparable to those reported here. The main difference was in the variances of our estimated measures of fit. N = 5,000 was chosen because at that level the variance is much better than at N = 1,000 and not much worse than that at N = 10,000.

The basic estimators used to compare recovery of F from case to case were the empirical distribution function (EDF)

$$\tilde{F}_{N,J}(t) = \frac{1}{N}\sum_{n=1}^{N} 1_{\{\tilde{\theta}_J^{(1)}(\underline{X}_{nJ}) \le t\}}$$

and the kernel distribution estimator (KDE)

$$\hat{F}_{N,J}(t) = \frac{1}{N}\sum_{n=1}^{N} K\left[\frac{t - \tilde{\theta}_J^{(1)}(\underline{X}_{nJ})}{h}\right]$$

where

$$\tilde{\theta}_J^{(1)}(\underline{X}_J) = \overline{P}_J^{-1}\left[\frac{J \cdot \overline{X}_J + 1}{J+2}\right]$$

(and K and h are as described in (8) and (9) above). Each of these distribution estimators is consistent for the true Θ distribution, by application of Theorem 1 through Theorem 3.

For each simulated data set, sample means and standard deviations for estimates of

$$\text{RMS} = \left\{E\int_{-\infty}^{\infty}[F_{est}(t) - F(t)]^2 g(t)dt\right\}^{1/2}$$

are reported. In addition, mean estimates of

$$\text{MAX} = E[\sup\{|F_{est}(t) - F(t)| : -\infty \le t \le \infty\}]$$

and the average value LOC $= t_{\max}$ at which MAX is attained are reported. (Note: F_{est} stands for either of the distribution estimators above.) In general the weighting function g should be chosen to reflect our interests in the Θ distribution F: g should give more weight to areas of F that should be well-estimated and less weight to areas of F for which we are willing to tolerate less good estimation. In these simulations, the weighting function g was taken to be the standard normal density: some weight is given to estimating F well at all θ's, but more weight is given to estimating F well near $\theta = 0$. More details about these distances and the methods of calculation can be found in Appendix A below.

Test Length	Estimator	RMS		Deviation	
		ave	SD	MAX	LOC
10	EDF	0.04655	0.00002	0.11021	0.37694
	KDE	0.02318	0.00003	0.03812	0.89134
30	EDF	0.01692	0.00001	0.04032	0.09754
	KDE	0.00887	0.00002	0.01447	0.23184
60	EDF	0.00984	0.00002	0.02510	0.07844
	KDE	0.00652	0.00002	0.01076	0.05334
100	EDF	0.00731	0.00002	0.01895	-0.02856
	KDE	0.00577	0.00002	0.00965	-0.07616

Table 2: $\Theta \sim N(0,1)$, Rasch

Test Length	Estimator	RMS		Deviation	
		ave	SD	MAX	LOC
10	EDF	0.07015	0.00002	0.15724	-1.00076
	KDE	0.05158	0.00003	0.09368	-1.23646
30	EDF	0.02794	0.00002	0.06418	-0.77476
	KDE	0.02176	0.00002	0.03755	-1.26626
60	EDF	0.01521	0.00002	0.03527	-0.46316
	KDE	0.01251	0.00002	0.02109	-1.05756
100	EDF	0.01035	0.00002	0.02463	-0.33196
	KDE	0.00907	0.00003	0.01532	-0.80926

Table 3: $\Theta \sim N(0,1)$, 3PL

Test Length	Estimator	RMS		Deviation	
		ave	SD	MAX	LOC
10	EDF	0.09665	0.00004	0.22175	−0.74996
	KDE	0.08412	0.00004	0.13431	−1.21956
30	EDF	0.05695	0.00004	0.11573	−0.67436
	KDE	0.05439	0.00004	0.08258	−0.89616
60	EDF	0.01835	0.00002	0.04188	−0.70396
	KDE	0.01645	0.00003	0.02802	−1.10236
100	EDF	0.01823	0.00003	0.03782	−0.49826
	KDE	0.01767	0.00004	0.02668	−0.79636

Table 4: $\Theta \sim N(0,1)$, Estimated

From Tables 2, 3 and 4, it is clear that smoothing in the KDE is helping, especially with short tests. In comparing Tables 2 and 3 it is clear that the presence of the nonzero lower asymptote c is degrading the fits. This can be seen both in the reduced RMS values and in the movement of LOC, the location of the maximum deviation between F_{est} and F, toward negative values. Finally, comparison of Tables 3 and 4 indicates that using 'noisy' ICC's somewhat degrades the recovery of F.

Figure 1 illustrates the performance of the estimators in Table 3. The first three panels are probability-probability ($p-p$) plots of the estimated Θ distribution (vertical axis) against the true Θ distribution (horizontal axis), for 10, 30 and 60 items. Each panel depicts one of the 100 Monte Carlo trials for the corresponding line of Table 3. The step functions represent the EDF estimator and the smooth curve represents the KDE estimator. The closer each is to the solid diagonal line, the better the true probabilities of the Θ distribution are estimated. In particular for 30 or 60 items, estimated probabilities are quite close to true probabilities. The story is very similar for the performance of the estimators in Tables 2, 5 and 6 (see also Figure 3). The fourth panel in Figure 1 compares the density derived from the KDE estimator in panel three to with the true Θ density (some excessive bumpiness in the estimated density is attributable to the fact that the "window width" h was chosen to

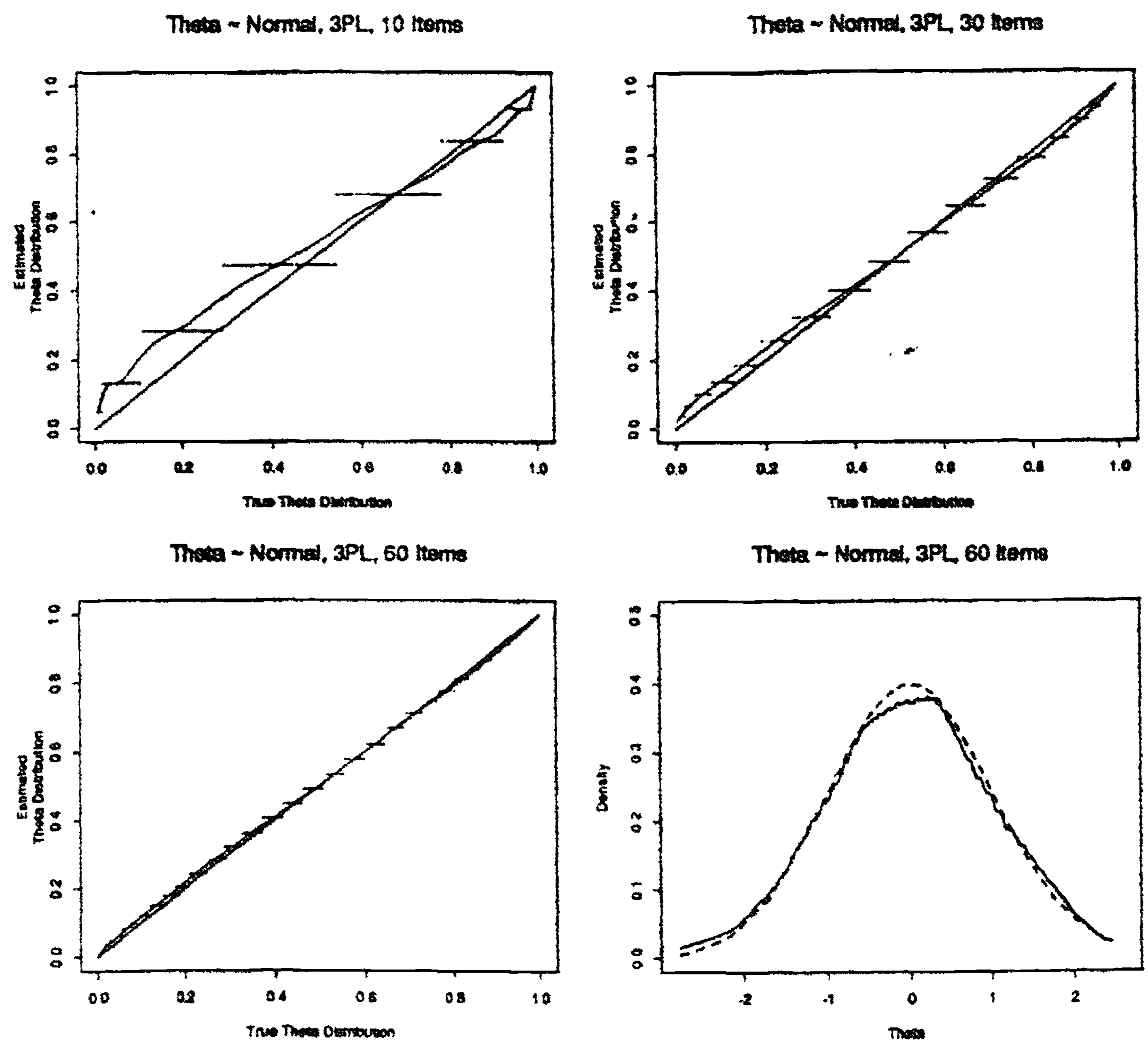

Figure 1: $p-p$ and density plots of EDF and KDE estimators. EDF is represented by step function, KDE by curve. In the last panel. the true density is the dashed curve and the KDE-based density estimate is the solid curve.

make a good ***distribution*** estimate rather than to make a good ***density*** estimate).

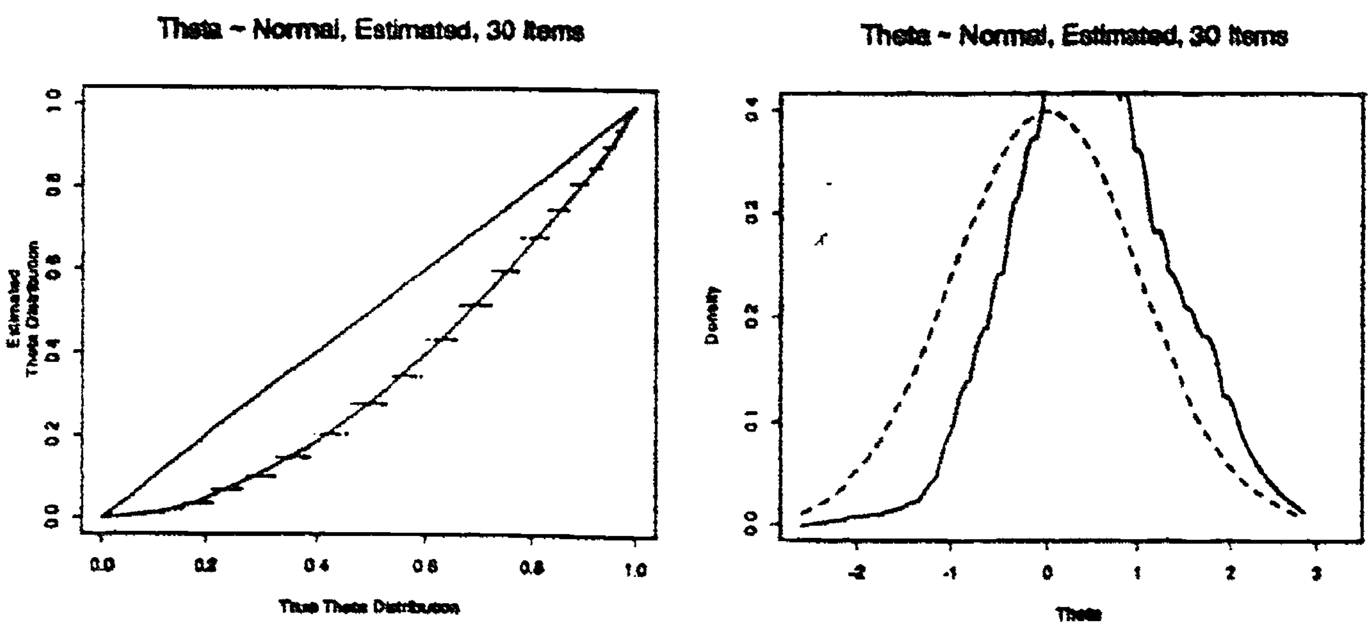

Figure 2: $p-p$ and density plots of EDF and KDE estimators. EDF is represented by step function, KDE by curve. In the second panel, the true density is the dashed curve and the KDE-based density estimate is the solid curve.

Figure 2 illustrates the performance of the estimators in Table 4. The left panel is a $p-p$ plot of the EDF (step function) and KDE (smooth curve) estimators for 30 items, and the right panel compares the corresponding KDE-based density with the true Θ density. In the Monte Carlo trial illustrated, contamination in the parameters of the "recovery" ICC's caused some bias and scale distortion in the estimated distribution, but the estimate still correctly suggests that Θ has a Normal or bell-shaped distribution.

In Tables 5, 6 and 7, in which Θ is bimodal, the KDE estimator is still doing better than the EDF. It is encouraging to see that the orders of magnitudes of the RMS and MAX measures of fit are the same as in the $N(0,1)$ case above. It is a little surprising that the fits can actually be better for the bimodal cases than the normal, but perhaps the greater variability is working in our favor here: we are getting more extreme-ability examinees with which to form F_{est} and thus to estimate the tails of F. Finally, note that there is much less difference in the fits of the 3PL and 'Estimated' 3PL cases.

Test Length	Estimator	RMS ave	RMS SD	Deviation MAX	Deviation LOC
10	EDF	0.04769	0.00003	0.12379	−1.36996
	KDE	0.03678	0.00003	0.06299	−1.25226
30	EDF	0.01820	0.00003	0.04668	−0.61856
	KDE	0.01547	0.00003	0.02502	−0.42646
60	EDF	0.01107	0.00003	0.02710	−0.25206
	KDE	0.00995	0.00003	0.01622	−0.17576
100	EDF	0.00870	0.00003	0.01923	−0.03886
	KDE	0.00817	0.00003	0.01290	−0.13216

Table 5: $\Theta \sim$ Bimodal, Rasch

Test Length	Estimator	RMS ave	RMS SD	Deviation MAX	Deviation LOC
10	EDF	0.05268	0.00003	0.12160	1.08084
	KDE	0.03612	0.00003	0.09342	−4.44996
30	EDF	0.02268	0.00002	0.05616	−0.66696
	KDE	0.01877	0.00002	0.04229	−3.68386
60	EDF	0.01353	0.00003	0.03496	−1.24996
	KDE	0.01205	0.00003	0.02561	−2.75386
100	EDF	0.00998	0.00003	0.02457	−1.22086
	KDE	0.00924	0.00003	0.01860	−2.64946

Table 6: $\Theta \sim$ Bimodal, 3PL

Figure 3 illustrates the performance of the estimators in Table 6, for 60 items. Again, the left panel is a $p-p$ plot of the EDF (step function) and KDE (smooth curve) estimators and the right panel depicts the KDE-based density estimate. Once again the estimated distribution provides good estimates of probabilities under the true distribution, and the corresponding density estimate tracks the two modes of the true Θ distribution reasonably well.

Test Length	Estimator	RMS ave	RMS SD	Deviation MAX	Deviation LOC
10	EDF	0.06387	0.00005	0.14624	0.78714
	KDE	0.05101	0.00005	0.09497	−4.97589
30	EDF	0.03203	0.00005	0.08038	−2.37405
	KDE	0.02958	0.00005	0.06457	−3.38695
60	EDF	0.01386	0.00003	0.03747	−1.11546
	KDE	0.01245	0.00003	0.02796	−2.63776
100	EDF	0.01120	0.00004	0.02776	−1.42786
	KDE	0.01055	0.00004	0.02134	−2.29616

Table 7: $\Theta \sim$ Bimodal, Estimated

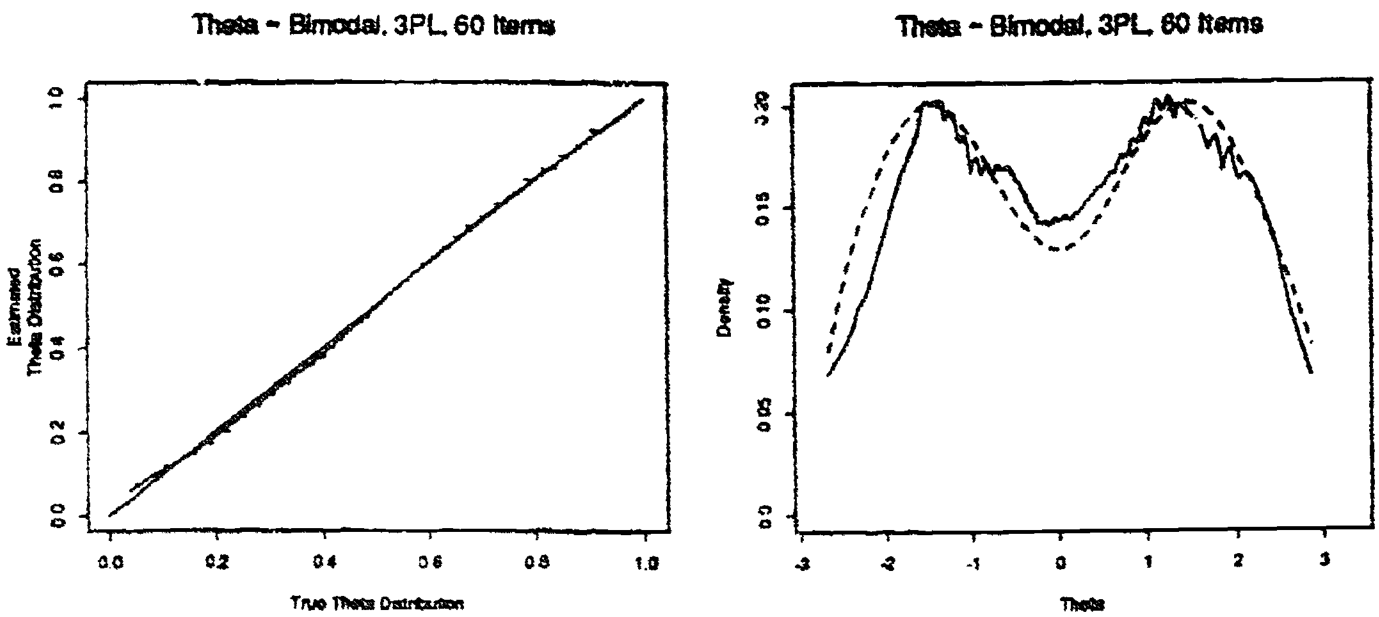

Figure 3: $p-p$ and density plots of EDF and KDE estimators. EDF is represented by step function, KDE by curve. In the second panel, the true density is the dashed curve and the KDE-based density estimate is the solid curve.

In Tables 8, 9 and 10, note how gradual the decrease in MAX is; this can be attributed partly to the fact that $\tilde{\theta}_J^{(1)}$ "doesn't know" that F assigns no mass to the interval $(-\infty, -1)$ and thus freely places $\tilde{\theta}$'s there. so that F_{est} is grossly overestimating F for $\theta < -1$. This certainly explains why LOC is near -1 in all but one case. It seems remarkable that the RMS should drop as much as it does, considering the fact that the Normal weighting function g assigns significant weight to the region near or below $\theta = -1$. Once again there is little difference between the 3PL and 'Estimated' 3PL cases. Finally, note that the EDF estimator is doing better than the KDE estimator in many cases here. Our ad hoc choice of h is probably failing us here by being too large to track the "sharp upturn" in F at -1.

Test Length	Estimator	RMS ave	RMS SD	Deviation MAX	Deviation LOC
10	EDF	0.09922	0.00004	0.23352	-0.26996
	KDE	0.09241	0.00003	0.20600	-1.00996
30	EDF	0.05404	0.00003	0.14608	-0.91796
	KDE	0.05508	0.00003	0.17924	-1.00996
60	EDF	0.03812	0.00003	0.15993	-1.00996
	KDE	0.04010	0.00003	0.16010	-1.00316
100	EDF	0.02944	0.00003	0.15246	-0.99996
	KDE	0.03215	0.00003	0.14717	-0.99996

Table 8: $\Theta \sim \chi^2 - 1$, Rasch

5 Discussion

To implement this scheme in practice, one must numerically invert the average ICC $\overline{P}_J$ for the test in question at or near the J+1 possible values of $\overline{X}_J$. After this, a table constructed from the inversion can be used simply and cheaply to estimate Θ distributions for each of several administrations of the same test, or each of several subpopulations in a single administration. For shorter tests lengths the basic statistic $\tilde{\theta}_J$ may need to be rescaled,

Test Length	Estimator	RMS ave	RMS SD	Deviation MAX	Deviation LOC
10	EDF	0.11871	0.00004	0.30689	–1.00996
	KDE	0.10699	0.00004	0.28934	–1.00996
30	EDF	0.07276	0.00004	0.22700	–1.00996
	KDE	0.07188	0.00004	0.23167	–1.00996
60	EDF	0.05291	0.00003	0.20477	–1.00996
	KDE	0.05408	0.00003	0.20211	–1.00996
100	EDF	0.04153	0.00003	0.19628	–0.99996
	KDE	0.04365	0.00003	0.18294	–1.00976

Table 9: $\Theta \sim \chi^2 - 1$, 3PL

Test Length	Estimator	RMS ave	RMS SD	Deviation MAX	Deviation LOC
10	EDF	0.11387	0.00005	0.30689	–1.00996
	KDE	0.10600	0.00005	0.33073	–1.00996
30	EDF	0.08264	0.00005	0.32359	–1.00996
	KDE	0.08161	0.00005	0.30244	–1.00996
60	EDF	0.05322	0.00003	0.20477	–1.00996
	KDE	0.05466	0.00004	0.21590	–1.00996
100	EDF	0.04303	0.00004	0.20150	–1.00996
	KDE	0.04491	0.00004	0.20859	–1.00646

Table 10: $\Theta \sim \chi^2 - 1$, Estimated

as we have done with $\tilde{\theta}_J^{(1)}$, to effectively estimate F. Kernel smoothing of the estimated distribution (KDE) is also quite helpful. Work is currently underway (Nandakumar and Junker, 1992) to further examine and refine these methods using essentially unidimensional simulation data, and to apply the estimators to real tests.

Because it is fast, this scheme could be also be used for some diagnostic purposes. For example, if ICC's were estimated under the assumption of a Normal underlying Θ distribution and a 3PL model, the KDE estimate of the Θ distribution could be plotted on a Normal probability plot to examine (jointly) the assumptions about distribution and ICC forms. Or the Θ distribution estimates under two ICC estimation techniques could be compared to see how well they agree: Quite different ICC forms or parameter sets could in principle lead to very similar Θ distributions; if so then for many purposes it would then be a matter of indifference which ICC's were used, so considerations such as cost of ICC estimation, etc., could come into play. Finally, it may be possible to estimate the Θ distribution sufficiently accurately with, say, Rasch ICC's (for which item parameters can be estimated independently of the Θ distribution), and then use that estimate as part of a marginal maximum likelihood approach to estimating item parameters in a 3PL model which more accurately models the item response behavior.

References

Ash, R. (1972). *Real Analysis and Probability.* New York: Academic Press.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability, in Lord, F. and Novick, M. (1968). *Statistical Theories of Mental Test Scores.* Reading, Mass: Addison-Wesley.

Hambleton, R. K. (1989). Principles and selected applications of item response theory. In Linn, R.L. (ed.) *Educational Measurment, Third Edition.* New York: MacMillan, pp.

147–200.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the Americ Statistical Association*, *58*, 13–30.

Junker, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika*, *56*, 255–278.

Levine, M. (1984). *An introduction to multilinear formula score theory.* University of Illinois and Office of Naval Research, Research Report 84-4.

Lord, F. (1953). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, *13*, 517–549.

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

Nandakumar, R. and Junker. B. W. (1992). Estimating the latent ability distribution. To be presented at the Annual Meeting of the Psychometric Society, Columbus OH, July 1992.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method Method.* New York: John Wiley and Sons.

Samejima, F. (1984). *Plausibility functions of Iowa Vocabulary Test items estimated by the simple sum procedure of the conditional P.D.F. approach.* University of Tennessee and Office of Naval Research. Research Report 84-1.

Samejima, F. and Livingston. P. (1979). *Method of moments as the least squares solution for fitting a polynomial.* University of Tennessee and Office of Naval Research, Research Report 79-2.

Serfling, R. (1980). *Approximation Theorems in Mathematical Statistics.* New York: John Wiley and Sons.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis* London: Oxford University Press.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*, 589–617.

Stout, W. F. (1990). A new item response theory modeling approach with applications to undimensionality assessment and ability estimation. *Psychometrika, 55*, 293–325.

Appendix A Details of the simulation

For each simulated data set, M Monte Carlo trials were run (one trial entails sampling N examinees, generating a θ and J item responses for each examinee, and constructing the distribution estimates $\bar{F}_{N,J}$ and $\hat{F}_{NJh}$ from these). In our simulation, M was taken to be 100. In the discussion below, F_{est} stands for either of the two distribution estimates tried.

For each trial, two measures of fit to the true ability distribution F were reported. First, the value of

$$\tilde{S} = \max\{|F_{est}(t_1) - F(t_1)| : t_0, \ldots, t_{1200}\}$$

was calculated, for t_i's ranging from −6 to 6 spaced at 0.01 intervals, as an approximation to

$$S = \sup\{|F_{est}(t) - F(t)|; t \in (-\infty, \infty)\}$$

as well as the value $\tilde{L} = t_{i_{max}}$ at which $\tilde{S}$ was attained. Second, an approximation to the squared distance

$$I^2 = \int_{-\infty}^{\infty} [F_{est}(t) - F(t)]^2 g(t) dt$$

was calculated, where the weight function g was taken to be the standard normal density. The approximation used was the Monte Carlo approximation

$$\tilde{I}^2 = \frac{1}{K} \sum_{k=1}^{K} [F_{est}(T_k) - F(T_k)]^2,$$

where $T_1, \ldots T_K$ are iid with marginal density g, and $K = 500$ for our simulation.

Finally, Monte Carlo sample averages

$$\overline{S} = \frac{1}{M}\sum_{m=1}^{M} \tilde{S}_m,\ \overline{L} = \frac{1}{M}\sum_{m=1}^{M} \tilde{L}_m,\ \text{and}\ I^2 = \frac{1}{M}\sum_{m=1}^{M} \tilde{I}^2_m$$

were computed, as well as sample standard deviations. $\overline{S}$ estimates $E[\overline{S}]$, $\overline{L}$ estimates E[L], and $\overline{L}$ estimates $\{E[\overline{I}^2]\}^{1/2}$ standard deviation for $\overline{I}$ was estimated using the delta method (Serfling, 1980, p. 118).

$E[\overline{S}]$ may be regarded as a reasonable approximation to $MAX = E[S]$. Because of the discretization in calculating $\tilde{S}$ and $\tilde{L}$, $E[\overline{L}]$ probably is not as good an indication of the true value LOC $= t$ where the distributions are farthest apart, but it may still be of some descriptive value. Finally, $\{E[\overline{I}^2]\}^{1/2}$ is exactly

$$\text{RMS} = \left\{ E\int_{-\infty}^{\infty} [F_{est}(t) - F(t)]^2 g(t)dt \right\}^{1/2}$$

The pseudo-random number generators used were linear congruential generators (see Rubinstein, 1981)

$$r_\nu = (a \cdot r_{\nu-1} + c) \bmod m,$$

using $a = 7^5, c = 0, m = 2^{31}$ for generating Θ's and $a = 2^7 + 1,\ c = 1,\ m = 2^{35}$ for generating item responses. Normal observations were obtained from these uniform observations by the polar transformation

$$\begin{aligned} Z_1 &= \sqrt{-2\log U_1}\cos 2\pi U_2 \\ Z_2 &= \sqrt{-2\log U_1}\sin 2\pi U_2 \end{aligned}$$

and the bimodal mixture and $\chi^2 - 1$ observations were taken to be appropriate transformations of these. Pseudo-random values obtained using these transformations do exhibit some lattice structure but this was not considered a problem for our calculations, which are essentially all Monte Carlo integrations.

Appendix B Proofs

Proof of Theorem 1: Observe that, for any $\epsilon > 0$,

$$\begin{aligned} P\left[|\tilde{F}_{N,J}(\Theta) - F(\Theta)| \geq \epsilon\right] &\leq P\left[|\tilde{F}_{N,J}(\Theta) - F_J(\Theta)| + |F_J(\Theta) - F(\Theta)| \geq \epsilon\right] \\ &\leq P\left[|\tilde{F}_{N,J}(\Theta) - F_J(\Theta)| \geq \epsilon/2\right] \text{ (for large } J) \\ &\leq C \cdot e^{-2N(\epsilon/2)^2} \end{aligned}$$

for some universal constant C, and N large. (Serfling, 1980, p. 59). This tends to zero as $N \to \infty$. □

Proof of Theorem 2: Observe that

$$\begin{aligned} P[\overline{R}_J^{-1}(\overline{X}_J) \leq \theta] &= P[\overline{X}_J \leq \overline{R}_J(\theta)] \\ &= P[\overline{P}_J^{-1}(\overline{X}_J) \leq \overline{P}_J^{-1}\overline{R}_J(\theta)] \\ &= P[\overline{P}_J^{-1}(\overline{X}_J) + \tau(\theta) - \overline{P}_J^{-1}\overline{R}_J(\theta) \leq \tau(\theta)]. \end{aligned}$$

By Slutsky's Theorem, since $\tau(\theta) = \lim_{J\to\infty} \overline{P}_J^{-1}\overline{R}_J(\theta)$ we know that $\overline{P}_J^{-1}(\overline{X}_J)+\tau(\theta)$ and $\overline{P}_J^{-1}(\overline{X}_J)$ have the same asymptotic law, i.e. for any t,

$$P[\overline{P}_J^{-1}(\overline{X}_J) + \tau(\theta) - \overline{P}_J^{-1}\overline{R}_J(\theta) \leq t] \to F(t).$$

Then in particular for $t = \tau(\theta)$,

$$P[\overline{P}_J^{-1}(\overline{X}_J) + \tau(\theta) - \overline{P}_J(\theta)\overline{R}_J(\theta) \leq \tau(\theta)] \to F(\tau(\theta)).$$

The assertion about uniform convergence follows from a theorem of Polya (Serfling, 1980, p.18). □

Proof of Theorem 3: In the following calculation, it will be helpful to let Y be a random variable with distribution K independent of Θ and all item responses. Squaring (6),

$$\begin{aligned} RMS^2 &= E\int_{-\infty}^{\infty} [\hat{F}_{NJh}(t) - F(t)]^2 g(t)dt \\ &= \int_{-\infty}^{\infty} E\left\{\sum_{j=0}^{J} \hat{P}_N[\overline{X}_J = j/J]K\left[\frac{t - \overline{P}_J^{-1}(j/J)}{h}\right] - P[\Theta \leq t]\right\}^2 g(t)dt \end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \left\{ [bias(t)]^2 + [variance(t)] \right\} g(t)dt \\
&= \int_{-\infty}^{\infty} \left\{ \sum_{j=0}^{J} P_N \left[\overline{X}_J = j/J \right] K \left[\frac{t - \overline{P}_J^{-1}(j/J)}{h} \right] - P[\Theta \leq t] \right\}^2 g(t)dt \\
&\quad + \int_{-\infty}^{\infty} \operatorname{Var} \left\{ \sum_{j=0}^{J} \hat{P}_N[\overline{X}_J = j/J] K \left[\frac{t - \overline{P}_J^{-1}(j/J)}{h} \right] \right\} g(t)dt \\
&= \int_{-\infty}^{\infty} \left\{ P[\overline{P}_J^{-1}(\overline{X}_J) + hY \leq t] - P[\Theta \leq t] \right\}^2 g(t)dt \\
&\quad + \int_{-\infty}^{\infty} \operatorname{Var} \left\{ \frac{1}{N} \sum_{n=1}^{N} K \left[\frac{t - \overline{P}_J^{-1}(\overline{X}_{nJ})}{h} \right] \right\} g(t)dt \\
&= \int_{-\infty}^{\infty} \{ P[\overline{P}_J^{-1}(\overline{X}_J) + hY \leq t] - P[\Theta \leq t] \}^2 g(t)dt \\
&\quad + \frac{1}{N} \int_{-\infty}^{\infty} \operatorname{Var} K \left[\frac{t - \overline{P}_J^{-1}(\overline{X}_J)}{h} \right] g(t)dt \\
&= (bias)_{NJh}^2 + (variance)_{NJh}.
\end{aligned}
$$

Note that $(bias)_{NJh}^2$ does not depend on N. As long as

$$E|Y| = \int |u| K(u)du < \infty,$$

we will have $hY \to 0$ in probability, so that by Slutsky's Theorem the distributions of $\overline{P}_J^{-1}(\overline{X}_J) + hY$ and $\overline{P}_J^{-1}(\overline{X}_J)$ will converge to the same thing, namely $F(t) = P[\Theta \leq t]$, at every t (we are assuming F is continuous) as $J \to \infty$ and $h \to \infty$ and $h \to 0$. Hence the integrand of $(bias)_{NJh}^2$ converges to zero at each t, and if $g(t)$ is a density it follows that $(bias)_{NJh}^2 \to 0$ as $J \to \infty$ and $h \to 0$ (and N is free).

On the other hand, for each fixed J, h, t the random variable

$$K \left[\frac{t - \overline{P}_J^{-1}(\overline{X}_J)}{h} \right]$$

is bounded between 0 and 1. hence if $g(t)$ is a density we have for each fixed J and h

$$\int_{-\infty}^{\infty} \operatorname{Var} K \left[\frac{t - \overline{P}_J^{-1}(\overline{X}_J)}{h} \right] g(t)dt < 1.$$

Multiplying by $1/N$ it is clear that $(variance)_{NJh} \to 0$ as $N \to \infty$ uniformly in J and h. This proves Theorem 3. □

Dr. Terry Ackerman
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Ronald Armstrong
Rutgers University
Graduate School of Management
Newark, NJ 07102

Dr. William M. Bart
University of Minnesota
Dept. of Educ. Psychology
330 Burton Hall
178 Pillsbury Dr., S.E.
Minneapolis, MN 55455

Dr. Bruce Bloxom
Defense Manpower Data Center
99 Pacific St.
Suite 155A
Monterey, CA 93943-3231

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

Dr. John M. Carroll
IBM Watson Research Center
User Interface Institute
P.O. Box 704
Yorktown Heights, NY 10598

Mr. Hua Hua Chung
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

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

Dr. Timothy Davey
American College Testing Program
P.O. Box 168
Iowa City, IA 52243

Dr. Lou DiBello
CERL
University of Illinois
103 South Mathews Avenue
Urbana, IL 61801

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

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

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

Dr. Isaac Bejar
Law School Admissions
Services
P.O. Box 40
Newtown, PA 18940-0040

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

Dr. Gregory Candell
CTB/McGraw-Hill
2500 Garden Road
Monterey, CA 93940

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

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

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

Dr. C. M. Dayton
Department of Measurement
Statistics & Evaluation
College of Education
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

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314

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

Dr. Laura L. Barnes
College of Education
University of Toledo
2801 W. Bancroft Street
Toledo, OH 43606

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

Dr. Robert Breaux
Code 281
Naval Training Systems Center
Orlando, FL 32826-3224

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

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

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Ms. Carolyn R. Crone
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

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

Mr. Hei-Ki Dong
Bell Communications Research
Room PYA-IK207
P.O. Box 1320
Piscataway, NJ 08855-1320

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

ERIC Facility-Acquisitions
2440 Research Blvd. Suite 550
Rockville, MD 20850-3238

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

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

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

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

Dr. Sherrie Gott
AFHRL/MOMJ
Brooks AFB, TX 78235-5601

Prof. Edward Haertel
School of Education
Stanford University
Stanford, CA 94305

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

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

Dr. William Howell
Chief Scientist
AFHRL/CA
Brooks AFB, TX 78235-5601

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

Dr. Benjamin A. Fairbank
Operational Technologies Corp.
5825 Callaghan, Suite 225
San Antonio, TX 78228

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

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

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

Dr. Drew Gitomer
Educational Testing Service
Princeton, NJ 08541

[illegible] Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

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

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

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

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

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

Dr. Marshall J. Farr, Consultant
Cognitive & Instructional Sciences
2520 North Vernon Street
Arlington, VA 22207

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

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

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

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

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

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

Dr. Thomas M. Hirsch
ACT
P. O. Box 168
Iowa City, IA 52243

Ms. Julia S. Hough
Cambridge University Press
40 West 20th Street
New York, NY 10011

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. Kumar Joag-dev
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign, IL 61820

Dr. Douglas H. Jones
1280 Woodfern Court
Toms River, NJ 08753

Dr. Brian Junker
Carnegie-Mellon University
Department of Statistics
Schenley Park
Pittsburgh, PA 15213

Dr. Michael Kaplan
Office of Basic Research
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

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. Jwa-keun Kim
Department of Psychology
Middle Tennessee State
University
P.O. Box 522
Murfreesboro, TN 37132

Dr. Soon-Hoon Kim
Kedi
92-6 Umyeon-Dong
Seocho-Gu
Seoul
SOUTH KOREA

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. Richard J. Koubek
Department of Biomedical
& Human Factors
139 Engineering & Math Bldg.
Wright State University
Dayton, OH 45435

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

Mr. Rodney Lim
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

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. Richard Luecht
ACT
P. O. Box 168
Iowa City, IA 52243

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. Clessen J. Martin
Office of Chief of Naval
Operations (OP 13 F)
Navy Annex, Room 2832
Washington, DC 20350

Dr. James R. McBride
HumRRO
6430 Elmhurst Drive
San Diego, CA 92120

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

Mr. Christopher McCusker
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Dr. Robert McKinley
Educational Testing Service
Princeton, NJ 08541

Mr. Alan Mead
c/o Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Chicago, IL 61801

Dr. Timothy Miller
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

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

Dr. Ratna Nandakumar
Educational Studies
Willard Hall, Room 213E
University of Delaware
Newark, DE 19716

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

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

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

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

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

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

Dr. Randall Shumaker
Naval Research Laboratory
Code 5510
4555 Overlook Avenue, S.W.
Washington, DC 20375-5000

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

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

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

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

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

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

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

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

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

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

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

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Headquarters Marine Corps
Code MPI-20
Washington, DC 20380

Librarian
Naval Center for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington, DC 20375-5000

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

Dr. Peter J. Pashley
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

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

Mr. Steve Reise
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

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

Dr. Mary Schratz
4100 Parkside
Carlsbad, CA 92008

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

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

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

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
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Dr. Maurice Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

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

Mr. Thomas J. Thomas
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

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. 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. Michael T. Waller
University of Wisconsin-Milwaukee
Educational Psychology Department
Box 413
Milwaukee, WI 53201

Dr. Ming-Mei Wang
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Dr. Thomas A. Warm
FAA Academy AAC934D
P.O. Box 25082
Oklahoma City, OK 73125

Dr. Brian Waters
HumRRO
1100 S. Washington
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
Federal Aviation Administration
800 Independence Ave, SW
Washington, DC 20591

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
02-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