DOCUMENT RESUME

ED 359 207 TM 019 833

AUTHOR Nandakumar, Ratna; Junker, Brian W.
TITLE Estimation of Latent Ability Distributions under Essential Unidimensionality.
PUB DATE Apr 93
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Ability; Elementary Secondary Education; *Estimation (Mathematics); Mathematical Models; Statistical Distributions; *Student Evaluation
IDENTIFIERS Ability Estimates; ACT Assessment; Empirical Distribution Function; Essential Independence (Tests); Item Characteristic Function; Kernel Method; *Latent Ability Distributions; Monitoring; Proportion Correct Score; Smoothing Methods; *Unidimensionality (Tests); Violation of Assumptions

ABSTRACT

In many large-scale educational assessments it is of interest to compare the distribution of latent abilities of different subpopulations, and track these distributions over time to monitor educational progress. B. Junker, together with two colleagues, has developed a simple scheme, based on the proportion correct score, for smoothly approximating the ability distribution from binary responses. The method works for essentially unidimensional models under essential independence. The smoothing parameter is refined, and the results obtained by Junker are replicated. The performance of the discrete empirical distribution function (EDF) and the kernel smoothed distribution estimate (KDE) in estimating ability distribution under essential unidimensionality was studied, and the methodology was illustrated with a real data set from 5,000 examinees taking the American College Test reading test. The KDE and EDF estimators are simple, fast, and easy to compute methods to recover latent distributions. These estimators work for a general class of item characteristic curves, and are robust under violations of local independence and strict unidimensionality assumptions. Seven tables and 34 graphs are included. (SLD)

Estimation of Latent Ability Distributions Under Essential Unidimensionality

Ratna Nandakumar
University of Delaware

and

Brian W. Junker
Carnegie Mellon University

Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA, April 13, 1993.

Estimation of Latent Ability Distributions Under Essential Unidimensionality

In many large–scale educational assessments (such as National Assessment of Educational Progress) it is of interest to compare the distribution of latent abilities of different subpopulations, and track these distributions over time to monitor educational progress. Several researchers have attempted to develop methodologies to recover ability distribution from item response data. For example, Samejima and Livingston (1979) have fit polynomials to latent densities using the method of moments. Samejima (1984) also fits Θ densities using MLE $\hat{\theta}$ by matching two or more moments. Levine (1984) projects the latent distribution onto a convenient function space and estimates projections by maximum likelihood methods. Mislevy (1984) adopts marginal maximum likelihood method to recover the distribution of the latent variable from the observed item response patterns.

All the above mentioned methods rely upon the assumption of local independence for their validity, and are computationally intensive. Junker (1988, 1992) in association with Paul Holland (ETS) and William Stout (Illinois), developed a simple scheme, based on the proportion correct score, for smoothly approximating the ability distribution from binary responses. His approach is also robust to some violations of local independence. Namely, the methodology works for essentially unidimensional models under essential independence.

Junker's Approach for Estimating the Latent Trait Distribution

Let J denote the number of binary items, $\underline{X}_J \equiv (X_1, X_2, \ldots, X_J)$, denote item response vector, and $P_1(\theta)$, $P_2(\theta), \ldots, P_J(\theta)$ denote the corresponding item characteristic curves ICC's with respect to Θ, where Θ denotes the latent trait of interest. Then, $\overline{X}_J = \frac{1}{J}\Sigma X_j$ is the average item score, and $\bar{P}_J(\theta) = \frac{1}{J}\Sigma P_j(\theta)$ is the average ICC

Under the usual assumptions of local independence (LI) and monotonicity (M), or more generally under Stout's (1987, 1990) formulation of essential independence (EI) and locally asymptotic discrimination (LAD), it can be shown that (Junker, 1988)

$$\tilde{\theta}_J(\underline{X}_J) = \bar{P}_J^{-1}(\bar{X}_J)$$

is a plausible point estimator of Θ. That is, $\tilde{\theta}_J(\underline{X}_J)$ is a consistent estimator of θ under either set of assumptions. The distribution of $\tilde{\theta}_J(\underline{X}_J)$ is then given by

$$\mathrm{F}_J(t) = \mathrm{P}[\tilde{\theta}_J(\underline{X}_J) \leq t] \qquad (1)$$

A natural estimator of the Θ distribution given in equation (1) is the "empirical" distribution of $\tilde{\theta}_J$'s obtained by administering the test $\underline{X}_J$ to N examinees resulting in N response vectors $\underline{X}_{1J}, \underline{X}_{2J}, \ldots, \underline{X}_{NJ}$ and the corresponding θ estimates $\tilde{\theta}_J(\underline{X}_{1J})$, $\tilde{\theta}_J(\underline{X}_{2J}), \ldots, \tilde{\theta}_J(\underline{X}_{NJ})$. The empirical distribution of $\tilde{\theta}_J$'s is given by

$$\tilde{\mathrm{F}}_{N,J}(\mathrm{t}) \equiv \frac{1}{N}\sum_{n=1}^{N} 1_{\{\tilde{\theta}_J(\underline{X}_{nJ}) \leq t\}} \qquad (2)$$

$$= \{\text{fraction of } \tilde{\theta}_J(\underline{X}_{nJ})\text{'s} \leq t\}$$

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

It has been shown that, if the distribution function $\mathrm{F}(t)$ is continuous, the empirical distribution function $\tilde{\mathrm{F}}_{NJ}(t)$ converges in probability to F at each t as both $J \to \infty$ and $N \to \infty$ (Junker, 1988, 1992).

Practical Limitations

In applications, if P_J has a lower asymptote, and if $\overline{X}_J \leq c$, then $P_J^{-1}(\overline{X}_J)$ is set to $-\infty$. Although the probability of this happening decreases to zero as J tends to infinity, it still does happen with some frequency when J is small. Therefore, we must be concerned with recovering the Θ distribution whenever $\overline{X}_J$'s fall below the lower asymptote c (and similarly for an asymptote d < 1). Two adjustments were made to overcome this problem.

(i) Replace the point estimator $\hat{\Theta}_J$ with

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

($\tilde{\Theta}_J^{(1)}$ also converges to Θ and is bounded if the asymptotes of $\overline{P}_J$ are 0 and 1)

This first adjustment takes care of pbar's when the lower asymptote is 0 and the upper asymptote is 1. As a result of this adjustment the variance of the estimated theta distribution shrinks slightly, but this shrinkage reduces as the test length increases.

(ii) The numerical inverter of the function $\overline{P}_J$ is written (on the computer) such that it finds the root of a linear extrapolation of $\overline{P}_J(t)=\overline{X}_J$ when $\overline{X}_J$ lies outside the asymptotes of $\overline{P}_J$.

The second adjustment takes care of adjustments that the first adjustment can not handle. For example, if the guessing parameter c>0, for cases where $\overline{X}$ < c, the numerical inverter approximates and assigns a finite value for $\tilde{\Theta}_J^{(1)}$. This adjustment also occurs less frequently as the test length grows.

Kernel Smoothing

The basic estimator presented in equation (1) can also be written as

$$\tilde{F}_{N,J}(t) = \sum_{j=0}^{J} \hat{P}_N[X_J = j/J] 1_{\{P_J^{-1}(j/J) \leq t\}}$$

$$= \sum_{j=0}^{J} \hat{P}_N[X_J = j/J] \tilde{K}\left[\frac{t - P_J^{-1}(j/J)}{h}\right] \qquad (3)$$

where $\hat{P}_N[.]$ is the estimator of the discrete distribution of X_J based on N observations. $\tilde{K}(u)$ is constant except for a jump from 0 to 1 at u=0, and h is any fixed positive number.

In cases where the distribution is a truly continuous one, the performance of $\tilde{F}_{N,J}$ in equation (2) can be improved by replacing the discrete function $\tilde{K}(u)$ with a continuous distribution function $K(u)$ which increases from 0 to 1 as u ranges from $-\infty$ to ∞. Let

$$\hat{F}_{NJh}(t) = \frac{1}{N} \sum_{n=1}^{N} K\left[\frac{t - P_J^{-1}(j/J)}{h}\right] \qquad (4)$$

denote the smoothed estimator obtained by replacing $\tilde{K}$ with K. In equation (4) h (window width) is a parameter of the smoothing function. If h is large the smoothing function increases slowly, and if h is near zero, the smoothing function is steeper.

A practical question is: given N and J, what is a reasonable choice for h so as to get best possible estimator of θ? The formula for h (Silverman, 1986 pp.45–48; Reiss, 1981) is given by

$$h = C\, J^{-1/5}\, (\mathrm{var}\ \Theta)^{1/2} \qquad (5)$$

where C is an unknown constant that may be determined experimentally.

The smoothing kernel $K(t)$ is given by

$$K(t) = \begin{cases} 0, & t < -1 \\ 1/4(3t - t^3 + 2), & |t| \leq 1 \\ 1, & t > 1 \end{cases}$$

Other smoothing kernels could be chosen besides K(t) shown here. The advantages of K(t) are: (a) it is easy and fast to compute, (b) it is conservative about the tails of the estimated distribution.

In Junker's (1988, 1992) study, C=1/3 was chosen and var(θ) was approximated by the interquartile range of a uniform distribution. Junker investigated the performance of both the discrete empirical distribution function (EDF) in equation (3), and the kernel (smoothed) distribution estimate (KDE) in equation (4) in a Monte Carlo experiment. The following parameters were varied: test length (10, 30, 60, 100), ability distributions (normal, bimodal, discontinuous), ICC for generation (Rasch, 3PL), ICC for recovery (Rasch, 3PL, and 3PL with noise introduced). Sample sizes of 5000 examinees were simulated in all cases.

Junker's results showed that both KDE and EDF were able to recover the Θ distribution very well in all cases, with KDE performing better than EDF, especially for short tests. As the test length increased, the distance measures, RMS, decreased, and the smoothness of the distribution and the density plots improved.

The goals of the present study were to refine the smoothing parameter h and replicate the results obtained by Junker (1992); to investigate the performance of EDF and KDE in estimating the ability distribution under essential unidimensionality as opposed to strict unidimensionality; and to illustrate the methodology on a real data set.

Refinement of the Smoothing Parameter h

The aim was to find the best method for estimating the variance calculation of Θ, and best value for C in equation (5). Three methods of estimating variance were considered:

1. The interquartile range of the uniform distribution (same as before) (V1)
2. the interquartile range estimated from the frequency distribution of the observed data (V2)
3. the direct estimation of the variance from the frequency distribution of the observed data (V3).

Four different values for C were considered: $C = 1, 1/2, 1/3, 1/4$.

In order to achieve the best combination of C and variance estimation a 3x4 design was used.

Table A
Different combinations of C and variance for computing h

	C1=1	C2=1/2	C3=1/3	C4=1/4
V1	*C1V1**	*C2V1**	*C3V1**	*C4V1*
V2	*C1V2**	*C2V2**	*C3V3*	*C4V4*
V3	*C1V3**	*C2V3*	*C3V3*	*C4V4*

For each cell of Table A we studied the performance of the smoothed estimator KDE and the unsmoothed estimator EDF by varying the following parameters. Two ability distributions to generate θ's: Normal, bimodal; two types of ICC generation: 1PL, 3PL; two types of ICC's to recover the θ–distribution: 1PL, noisy 3PL (item parameters were deliberately contaminated with noise), two test lengths: 20, 60, and one examinee sample size: 1000. For each of the combinations of these parameters the performance of KDE and EDF was studied for those combinations of C and V marked with * in table A (the other cells did not produce promising results in initial simulations and therefore they were not studied further). These results are shown in Tables 1 to 4 and are based on 100 replications. In Tables 1 to 4, RMS EDF denotes the root mean square distance measure between the estimated and the true distributions for EDF estimator over 500 points averaged over replications. Similarly RMS KDE denotes the root mean square distance for the smoothed estimator. STD θ denotes the estimated standard deviation of the θ's averaged over replications.

Tables 1 and 2 show the results for the normal distribution and Tables 3 and 4 show the results for the bimodal distribution. In each of the tables, the column under RMS EDF across conditions is purely random and is not affected by the C and V combination. However, it shows how these values bounce around over replications. The column under RMS KDE shows the real differences in RMS's for different conditions. For example, in Table 1 RMS EDF for C1V3 is much smaller than for C3V1 with 20 items.

Figures 1–4 show the distribution and density plots for a sample of runs in Tables 1–4. Each of the figures contains two panels. The first panel to left is the P–P plot, where the X–axis denotes the true distribution and the Y–axis denotes the estimated distribution. The step function denotes the EDF estimator and the smooth curve denotes the KDE estimator. The closer each of the estimators to the solid diagonal line, the better the estimator. The right panel compares the density derived from the KDE estimator with the true density.

The decision to choose appropriate values for V and C were based on the distance measures in Tables 1–4 and plots of the estimated distribution function in Figures 1–4. Based on these results C1V3 combination was chosen. That is, the direct variance estimation method with the constant $C=1$ produced the best smoothing parameter h.

With the new values for C and V we then repeated the study done by Junker (1992) to see if we get similar results. These results are shown in Tables 5 and 6. Table 5 shows results for the normal distribution across three types of icc generation and recovery, and the Table 6 shows the same for the bimodal distribution. A sample of p–p plots and density plots are shown in Figures 5 and 6.

In Table 5 comparison of results across three types of ICC's shows that the distance measures slightly increase as the model gets more complex. That is, with guessing and noise present the error slightly increases. We also studied the location on the θ–scale where the maximum discrepancy occurs between the true and the estimated distributions. Across the models this location shifts to the left. In other words, as the guessing and noise are introduced, there is more instability in estimation at lower ability levels, as can be expected. The same findings were observed for bimodal distribution in Table 6. These results indicate that RMS and p–p plots are similar to or slightly better than those obtained by Junker (1992).

Essentially Unidimensional Study

The main goal of this study was to see if the results observed so far for the data generated with strictly unidimensional models would hold up for data generated with essentially unidimensional models, where there is one dominant dimension influencing responses to all items, and several minor dimensions influencing responses to a few items. The data generated here resembled a paragraph comprehension test where the dominant ability influenced all items and items of each paragraph were influenced by an additional ability unique to the paragraph. In all there were $1+r$ abilities, where r denotes the number of paragraphs. Each item is therefore influenced by two abilities, the dominant ability θ and one of the minor abilities, θ_1 through θ_r. The abilities were generated from a bivariate normal distribution with zero correlation between the abilities. The item parameters were generated as follows.

$$a_1 \sim N((1-\xi)\mu, \sqrt{1-\xi}\,\sigma)$$
$$a_2 \sim N(\xi\mu, \sqrt{\xi}\,\sigma)$$
$$b_i \sim N(0,1),\ i=1,2$$

where ξ denotes the strength of the minor ability in relation to the major ability.

The two–dimensional 3PL model was used to generate item responses, given by

$$P_i(\theta_1,\theta_2) = c_i + \frac{1-c_i}{1 + \exp\{-1.7\underline{a}_i'(\underline{\theta}-\underline{b})\}},$$

where $\underline{a}$ is the discrimination vector, $\underline{\theta}$ is the ability vector, and $\underline{b}$ is the difficulty vector of item i. Three test lengths were used (20, 40, 60), and two ξ values (0.2 and 0.4) were used in simulations. For each test length, 2items per paragraph and 5 items per paragraph were considered.

Two ways of estimating (recovering) ICCs were considered. First, the two sets of item parameters used for generating the item responses were manipulated as follows in order to obtain one parameter ICCs:

$$a(i) = \frac{a1(i)}{\sqrt{1+a2(i)*a2(i)}}$$

$$b(i) = b1(i) + (a2(i)/a1(i))\ b2(i),$$

which were then used to obtain the EDF and the KDE estimators. This definition of a(i) and b(i) is exactly what we would get if we averaged over the nuisance (paragraph) traits in the two–dimensional compensatory normal ogive model. This is a good approximation to the result of averaging over the nuisance traits in the logistic model we are using.

The results of this study are reported in Table 7 and the plots are shown in Figures 7–10. The RMS are within the expected range and the plots resemble those in strict dimensionality case. Hence one can conclude based on these simulations that the KDE estimator is an acceptable methodology to estimate the underlying distributions provided the ICCs can be well estimated.

Secondly, in order to investigate a more practical approach for estimating the ICCs, item parameters were obtained by using the computer program BILOG. These ICCs were then used to obtain the EDF and the KDE estimators. The results of this study are shown in Figures 11–16.

Figures 11–13 display distribution and density plots for the case where ξ=0.2 and Figures 14–16 display the plots for the case where ξ=0.4. That is, the influence of the minor ability in relation to the major ability is more in the later case. As can be seen from the figures, the distribution and densities are recovered smoothly in both cases. As the test length increases, the curves look smoother.

Real Data Study

In order to investigate the performance of KDE estimator on a real data set, ACT reading test was used. The reading test consists of 40 items and 4 paragraphs, where each paragraph is followed by 10 items. There were 5000 examinees is this data set. As a first step, DIMTEST (Stout, 1987; Nandakumar & Stout, 1993) was used to investigate if the reading test was essentially unidimensional. We found that the last 10 items were causing multidimensionality. Upon further investigation we found that the first 30 items tapped literature content area while the last 10 items tapped psychology content area. Moreover, since these were the last 10 items, speededness could have also caused the multidimensionality. When these items were removed, the rest of the items were found to be essentially unidimensional by DIMTEST (Nandakumar, in press). The item parameters of this data set were estimated using BILOG, and the ability distributions were estimated and compared using KDE estimators for three subpopulations: students who attained the grade A in high school (N=1574), students who attained the grade B (N=2144), and students who attained the grade C (N=915).

The comparison of distributions and densities for the three subpopulations are shown in Figure 17. From these plots it can be seen that the KDE estimator is smoothly estimating the distributions while the densities could be further improved. Notice that the estimated distribution for the A students is higher than those for the B and C students, and the estimated distribution for the B students is higher than that for the C students. This corresponds to our expectations, which helps to confirm the idea that the latent distribution estimator is performing well.

Summary and Discussion

In summary, the smoothing parameter h was refined to obtain optimal values for the constant C and for the variance estimation in equation (5) so that the KDE estimation of distributions are smooth. Secondly, performance of estimators EDF and KDE were investigated for the data generated under essentially unidimensional models. Two methods of estimating ICCs were considered for this purpose: (a) the two–dimensional item parameters used for generating the data were manipulated to resemble one–dimensional item parameters, (b) the ICCs were estimated by BILOG. In both cases RMSs, and the distribution and density plots indicated that these estimators are acceptable methods to estimate underlying ability distributions. Thirdly, the performance of the KDE estimator was illustrated on the ACT reading test to compare the distributions of three subpopulations. These results further confirmed that the KDE estimator is performing well to estimate latent distributions.

The KDE and EDF estimators investigated in this paper are simple, fast, and easy to compute methods to recover latent distributions. These estimators work for a general class of ICCs and are robust under violations of local independence and strict unidimensionality assumptions. The results of this paper illustrate promise of these methods for the future.

References

Junker, B. (1992). *A note on recovering the ability distribution from test scores.* (Tech. Rep No. N00014–91–J–1208). Pittsburgh: Carnegie Mellon University.

Junker, B. (1988). *Statistical aspects of a new latent trait model.* Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign.

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

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

Nandakumar, R. (in press). Assessing essential unidimensionality of real data. *Applied Psychological Measurement.*

Nandakumar, R. (1993). Refinements of Stout's procedure for assessing latent trait unidimensionality. *Journal of Educational Statistics, 18*, 41–68.

Reiss, R. D. (1981). Nonparametric estimation of smooth distribution functions. *Scandinavian Journal of Statistics, 8*, 116–119.

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., & 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.

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 unidimensionality assessment and ability estimation. *Psychometrika, 55*, 293–325.

Table 1: *Study:* Smoothing Parameter, $N = 1000$
θ–Distribution : Normal
ICC Generation : 1PL
ICC recovery : 1PL

Combination of C and V	20 Items			60 Items		
	RMS EDF	RMS KDE	STD θ	RMS EDF	RMS KDE	STD θ
C1V1	.0205	.0105	1.54	.0118	.0095	1.61
C1V2	.0217	.0045	.90	.0259	.0245	.94
C1V3	.0219	.0071	.98	.0192	.0184	.99
C2V1	.0218	.0075	1.55	.0131	.0118	1.61
C2V2	.0215	.0116	.91	.0194	.0174	1.00
C3V1	.0282	.0207	1.55	.0161	.0153	1.61

Table 2: *Study:* Smoothing Parameter, $N = 1000$
θ–Distribution : Normal
ICC Generation : 3PL
ICC recovery : Noisy 3PL

Combination of C and V	20 Items			60 Items		
	RMS EDF	RMS KDE	STD θ	RMS EDF	RMS KDE	STD θ
C1V1	.0339	.0329	1.70	.0155	.0158	1.80
C1V2	.0407	.0374	1.26	.0197	.0170	1.05
C1V3	.0362	.0318	1.40	.0122	.0077	1.02
C2V1	.0466	.0382	1.70	.0205	.0180	1.80
C2V2	.0430	.0375	1.26	.0136	.0100	.90
C3V1	.0385	.0325	1.70	.0128	.0091	1.80

Table 3: *Study:* Smoothing Parameter, $N = 1000$
θ–Distribution : ~~Normal~~ Bimodal
ICC Generation : 1PL
ICC recovery : 1PL

Combination of C and V	20 Items			60 Items		
	RMS EDF	RMS KDE	STD θ	RMS EDF	RMS KDE	STD θ
C1V1	.0274	.0245	1.55	.0239	.0232	1.61
C1V2	.0274	.0239	1.89	.0152	.0158	2.11
C1V3	.0286	.0263	1.54	.0134	.0130	1.68
C2V1	.0231	.0189	1.55	.0172	.0162	1.61
C2V2	.0221	.0183	2.07	.0111	.0100	2.18
C3V1	.0277	.0251	1.55	.0096	.0088	1.61

Table 4: *Study:* Smoothing Parameter, $N = 1000$
θ–Distribution : ~~Normal~~ Bimodal
ICC Generation : 3PL
ICC recovery : Noisy 3PL

Combination of C and V	20 Items			60 Items		
	RMS EDF	RMS KDE	STD θ	RMS EDF	RMS KDE	STD θ
C1V1	.0285	.0245	1.70	.0265	.0257	1.80
C1V2	.0383	.0365	2.17	.0141	.0134	2.10
C1V3	.0308	.0266	2.28	.0164	.0152	1.62
C2V1	.0324	.0295	1.70	.0180	.0169	1.80
C2V2	.0302	.0241	2.17	.0111	.0088	2.14
C3V1	.0264	.0215	1.70	.0150	.0139	1.80

Table 5: Replicatin Study with $C = 1$, $V = 3$, &, $N = 1000$.
θ - distribution – Normal.

Number of Items	ICC Generation – ICC recovery								
	1 PL – 1 PL			3 PL – 3PL			3 PL – 3 PL Nosiy		
	RMS EDF	RMS KDF	STD θ	RMS EDF	RMS KDF	STD θ	RMS EDF	RMS KDF	STD θ
20	:0229	.0107	.96	.0304	.0212	1.06	.0382	.0336	1.35
40	.0164	.0115	.97	.0197	.0157	1.04	.0334	.0319	1.08
60	.0147	.0119	.98	.0160	.0134	1.03	.0163	.0135	1.02

Table 6: Replicatin Study with $C = 1$, $V = 3$, &, $N = 1000$.
θ - distribution – Bimodal.

Number of Items	ICC Generation – ICC recovery								
	1 PL – 1 PL			3 PL – 3PL			3 PL – 3 PL Nosiy		
	RMS EDF	RMS KDF	STD θ	RMS EDF	RMS KDF	STD θ	RMS EDF	RMS KDF	STD θ
20	.0284	.0253	1.60	.0325	.0279	1.60	.0320	.0282	2.26
40	.0188	.0174	1.66	.0209	.0192	1.70	.0252	.0242	1.74
60	.0175	.0165	1.70	.0183	.0173	1.73	.0192	.0182	1.70

Table 7: Essential Unidimensionality:
Paragraph Comprehension
ICC Generation : Normal, two-dim 3 PL
ICC recovery : Normal, one-dim 3 PL

	$\xi = 0.2$, $d_E = 1$ 2 items/paragraph			$\xi = 0.2$, $d_E = 1$ 5 items/paragrah		
	RMS EDF	RMS KDE	STD θ	RMS EDF	RMS KDE	STD θ
20	.0455	.0379	1.03	.0276	[illegible]	1.00
40	.0292	.0245	1.06	.0256	.0238	1.12
60	.0150	.0109	1.01	.0131	.0094	1.02

Smoothing Parameter Study $d_L = 1$

Figure 1 C1V1 20 it

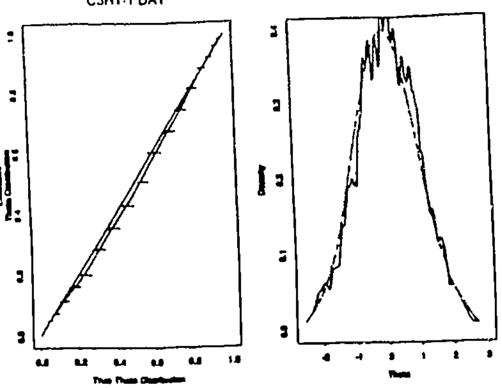


Figure 2 C2V1 30 it N = 30

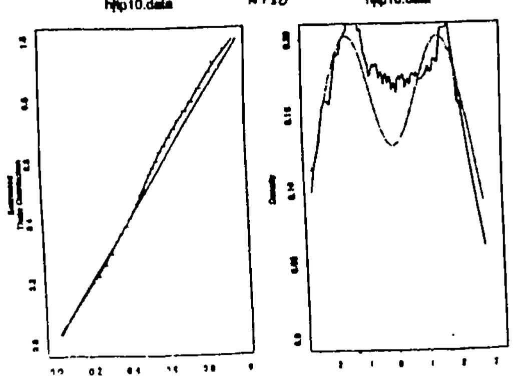


Figure 3 C1V3 20 it

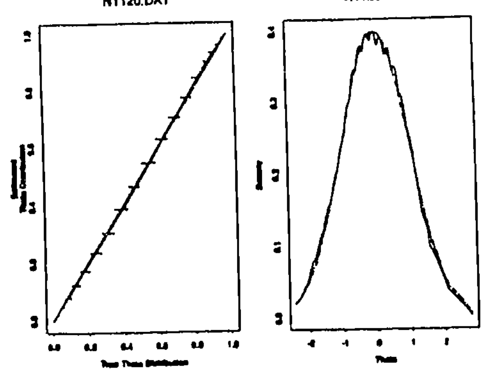
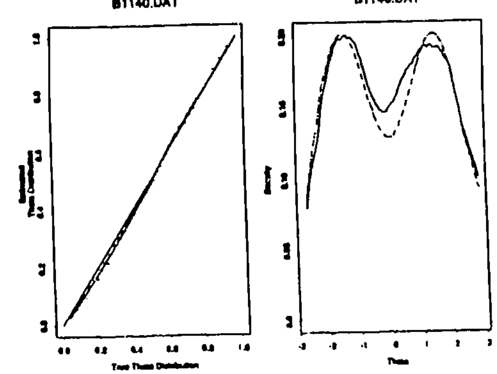


Figure 4 C1V3 40 it



Replication Study $d_L = 1$

Figure 5 C1V3 60 it

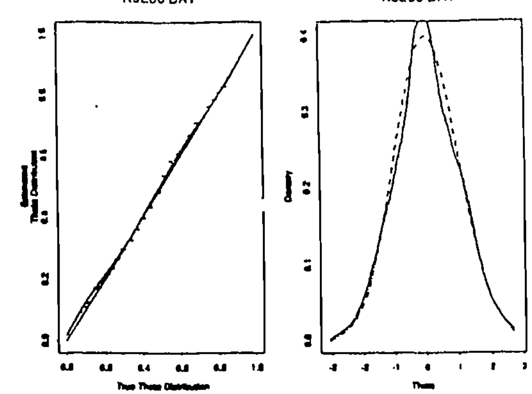
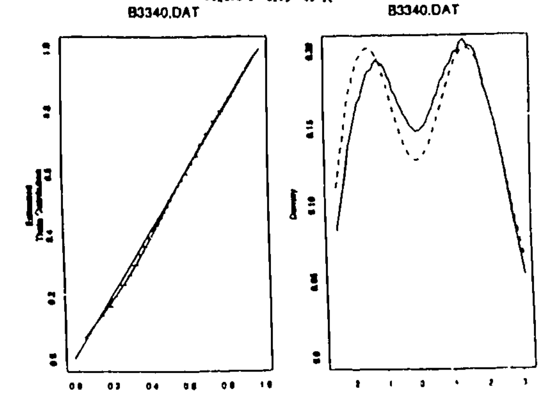


Figure 6 C1V3 40 it



18

Essential Unidimensionality Study $d_E = 1$

Fig. 7 Paragraph Comp. 20 it 2 it/para ξ = 0.2

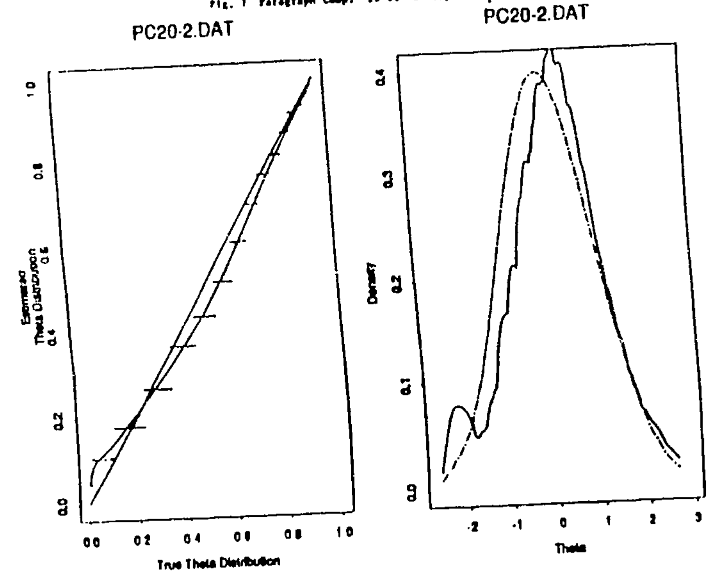


Fig. 8 Paragraph Comp. 40 it 5 it/para ξ = 0.2

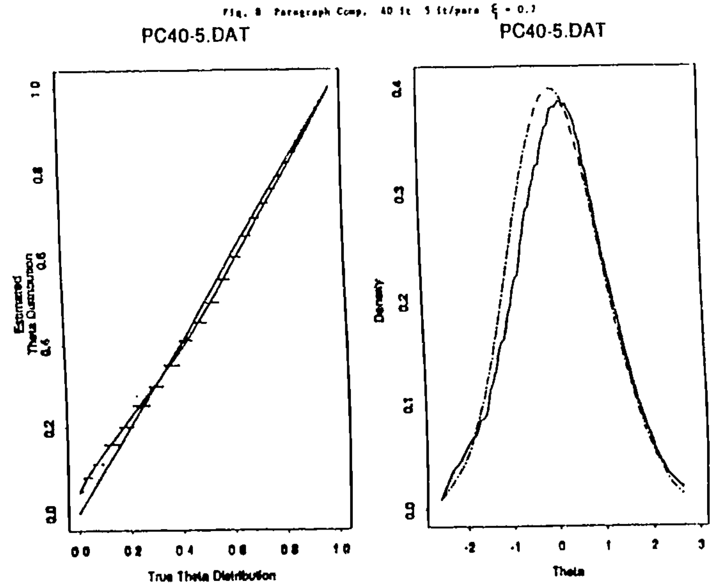


Fig. 9 Paragraph Comp. 50 it 5 it/para ξ = 0.2

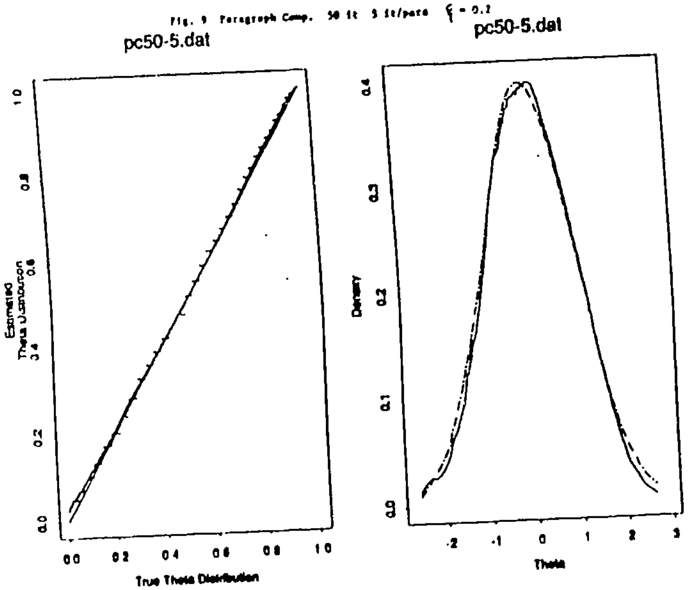
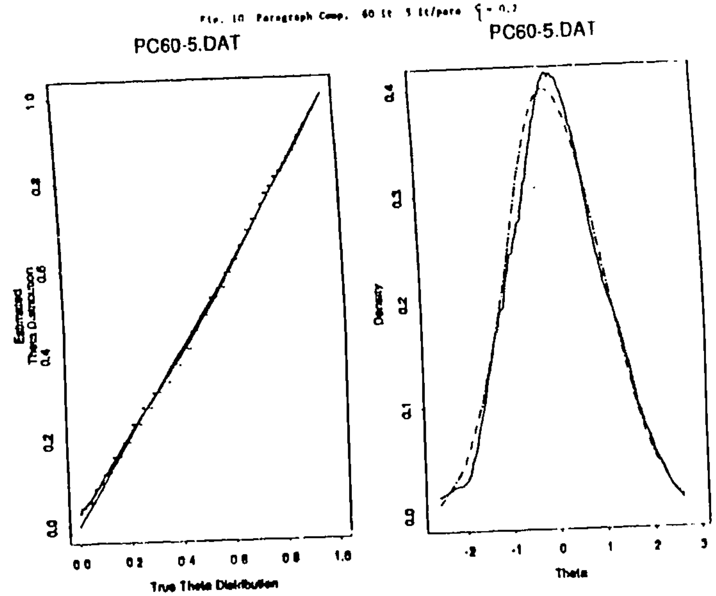


Fig. 10 Paragraph Comp. 60 it 5 it/para ξ = 0.2



Essential Unidimensionality Study = 0.2: Paragraph Comprehension Type
ICCS Estimated by BILOG

Figure 11 20 items

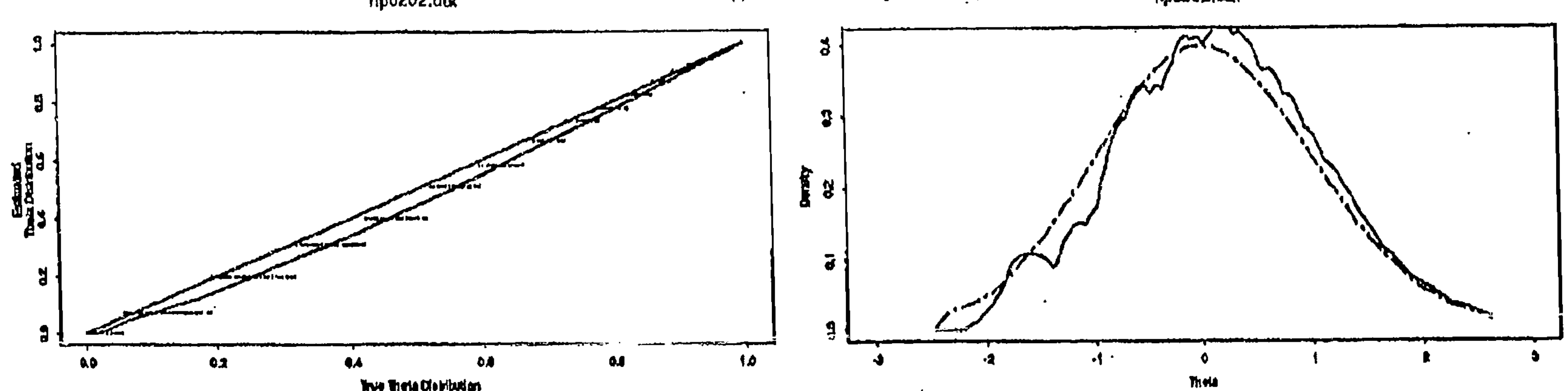


Figure 12 40 items

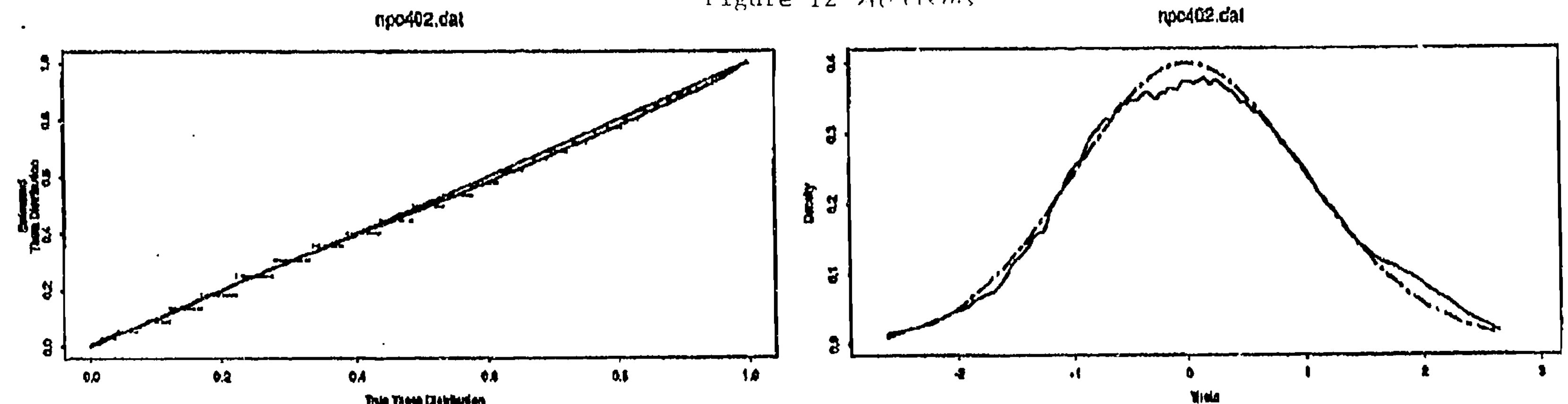
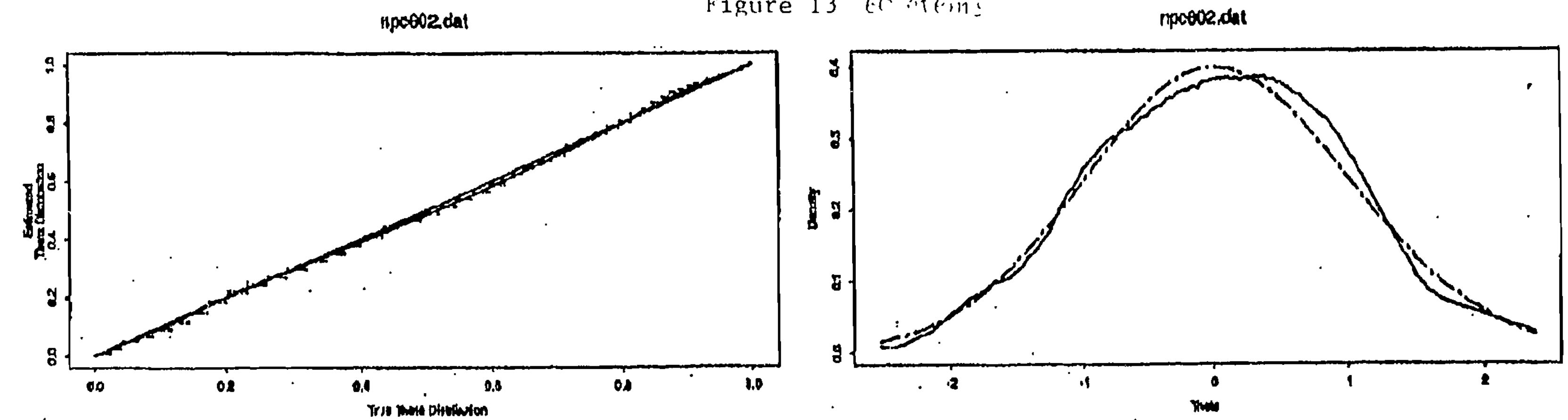


Figure 13 60 items



Essential Unidimensionality Study = 0.4: Paragraph Comprehension Type
ICCS Estimated by BILOG

Figure 14

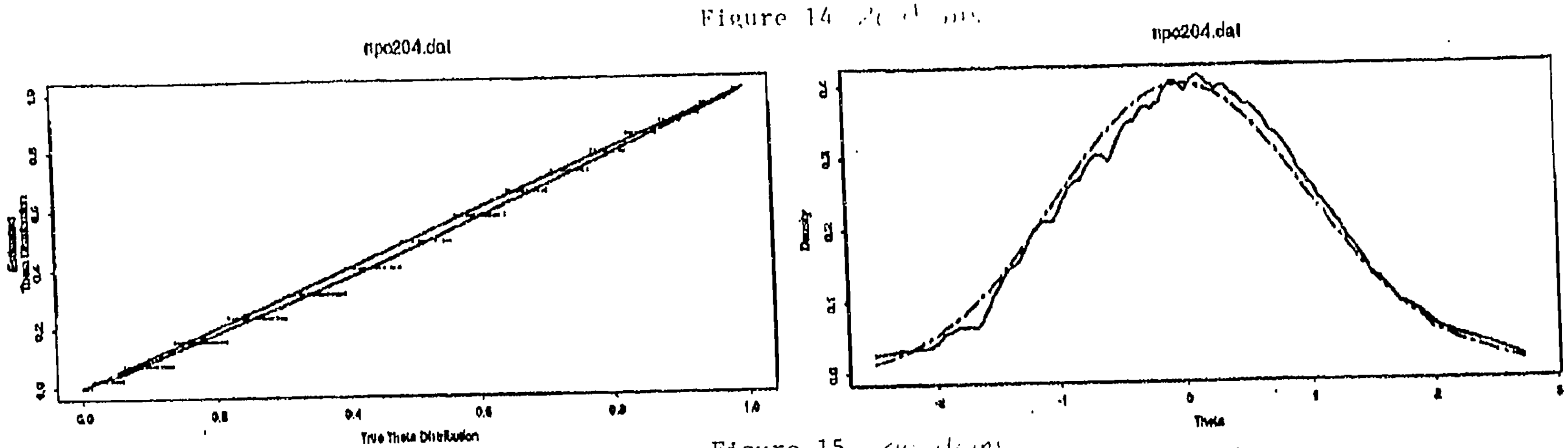


Figure 15

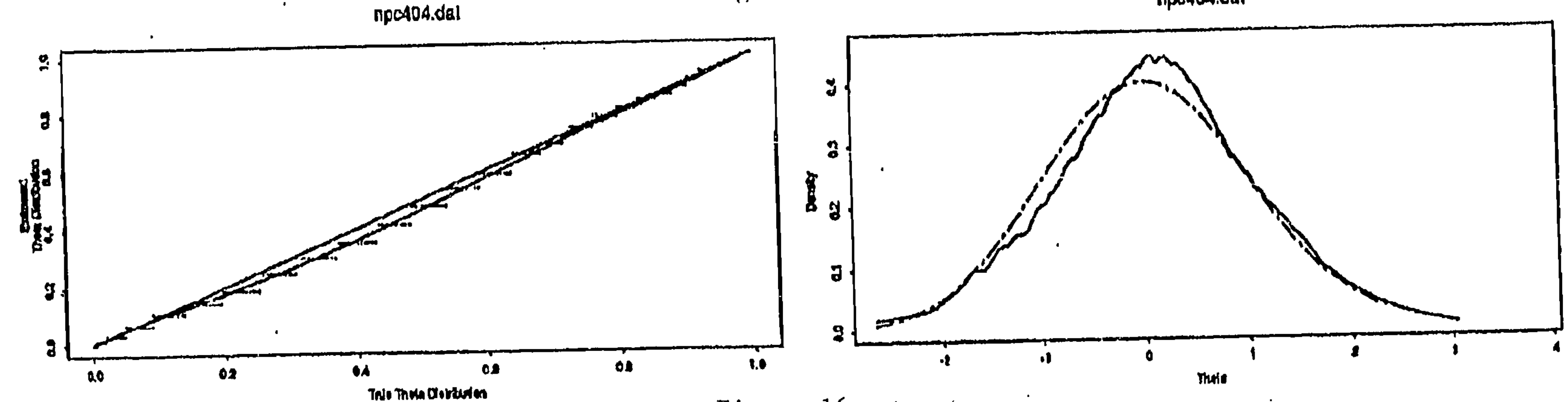


Figure 16

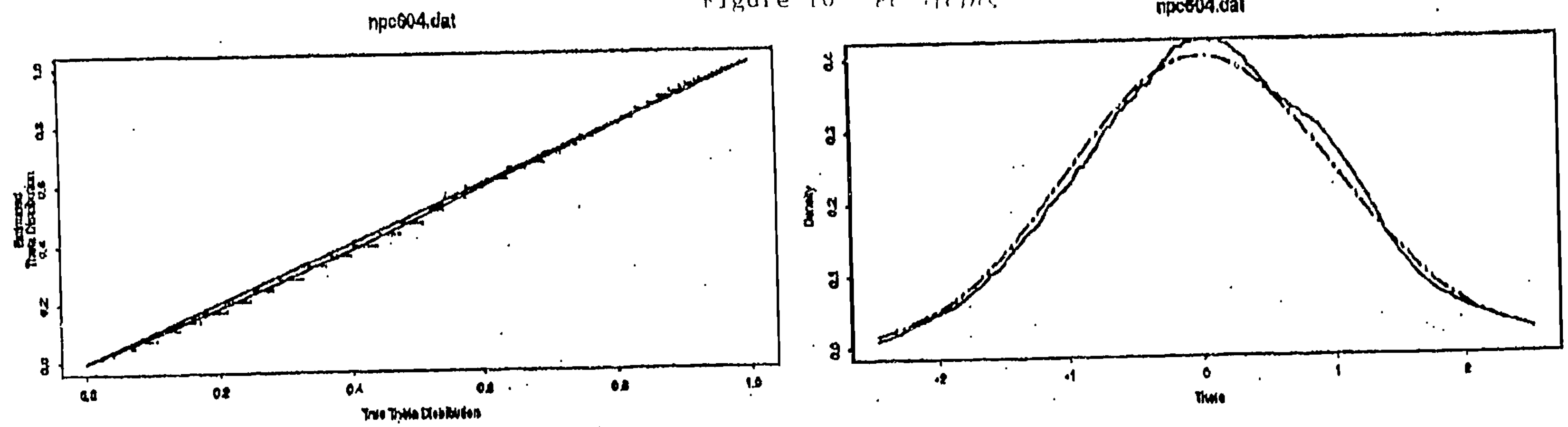


Figure 17: Reading Data Study
ICCS Estimated by BILOG