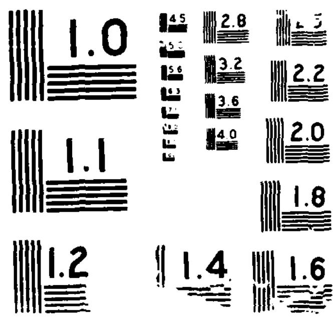
AD-A194 302 ADVANCEMENT OF LATENT TRAIT THEORY(U) TENNESSEE UNIV KNOXVILLE DEPT OF PSYCHOLOGY F SAMEJIMA FEB 88 N00014-81-C-0569 1/2

UNCLASSIFIED F/G 5/8 NL


1.0
1.1
1.2
1.4
1.6
1.8
2.0
2.2
2.8
3.2
3.6
4.0

(4)

ONR/FINAL REPORT

DTIC FILE COPY

AD-A194 302

ADVANCEMENT OF LATENT TRAIT THEORY

FUMIKO SAMEJIMA

UNIVERSITY OF TENNESSEE

KNOXVILLE, TENN. 37996-0900

FEBRUARY, 1988

Prepared under the contract number N00014-81-C-0569,
NR 150-467 with the
Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

R01-1068-71-003-88

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; Distribution unlimited.
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Fumiko Samejima, Ph.D Psychology Department 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION Cognitive Science 1142 CS
6c ADDRESS (City, State, and ZIP Code) 310 Austin Peay Building The University of Tennessee Knoxville, TN 37996-0900	7b ADDRESS (City, State, and ZIP Code) Office of Naval Research 800 N. Quincy Arlington, VA 22217
8a NAME OF FUNDING / SPONSORING ORGANIZATION Personnel and Training Research Programs 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-81-C-0569
8c ADDRESS (City, State, and ZIP Code) Office of Naval Research Arlington, VA 2217	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61153N	RR-042-04	RR 042-04-01	NR 150-467

11 TITLE (Include Security Classification)

Advancement of Latent Trait Theory

12 PERSONAL AUTHOR(S)

Fumiko Samejima, Ph.D

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final Report	FROM 1981 TO 1987	December 25, 1987	152

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Latent Trait Models, Maximum Likelihood Estimation, Bias, Cognitive Psychology, Continuous/Discrete Responses

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

This is a summary of the research conducted in the past six and half years, 1981-87, under the title, "Advancement of Latent Trait Theory."

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. Charles E. Davis	22b TELEPHONE (Include Area Code) 202-696-4741 22c OFFICE SYMBOL ONR-1142 CS

PREFACE

Almost six and half years have passed since I started this research on June 16, 1981, and these were hectic years. During this period, so many things were designed and accomplished. Even if I am the principal investigator, I find it practically impossible to include and systematize all the important findings and implications within a single final report, and it is my regret that so many of them have to be left out. I did my best within a limited amount of time, however, with the hope that this final report will help the reader to grasp the outline of the whole accomplishment.

There were six main objectives in the original research proposal, and they can be summarized as follows.

[1] Investigation of theory and method for estimating the operating characteristics of discrete item responses, which include the plausibility functions of the distractors of the multiple-choice test item, as well as the graded item responses of the free-response test item, without assuming any specific mathematical forms, and without using too many examinees in the whole procedure.

[2] Investigation of the various characteristics of the new family of models for the multiple-choice test item, both in theory and in practice.

[3] Production and revision of a set of systematic procedures for applying some combinations of a method and an approach for estimating the operating characteristics of discrete item responses, by modifying and reorganizing all the computer programs written for this purpose.

[4] Development of latent trait theory further, and include more varieties of situations.

[5] Investigation of ways of bridging across mathematical psychology and cognitive psychology, through latent trait theory.

[6] Systematizing theories and methods to eventually lead to a good introductory book on latent trait theory and other publications.

Out of these objectives, Objective [1] and [4], together with Objectives [3] and [5], were most intensively pursued. The highest productivity belongs to this part of the research. It provided us with valuable future perspectives of research. Objective [2] was also successfully pursued. In contrast to them, Objective [6] was more or less dropped. To compensate for it, however, some extensive research was done concerning the three-parameter logistic model. The main reason for this was because Navy had adopted the model for its computerized adaptive testing, and there was a necessity to pursue it.

It was my satisfaction and pleasure that Advanced Seminar on Latent Trait Theory was planned and held during this research period, and also that I had opportunities of introducing the research at international conferences as well as at domestic ones.

During the research period there were so many people who helped me as assistants, secretaries, etc., as I acknowledged in each research report. Also people of the Office of Naval Research, especially Dr. Charles E. Davis, and the people of the ONR Atlanta Office including Mr. Thomas Bryant and Mr.

Donald Calder, have been of great help in conducting the research. I would like to express my gratitude to all these people.

Thanks are also due to my two assistants, Christine A. Golik and Ali Khaddouma, and secretary, Betty Jo Allen, who helped me in preparing this final report. Appreciation is also extended to my former assistant Philip S. Livingston who still helped me occasionally during the research period.

December 25, 1987

Author

TABLE OF CONTENTS

I Introduction

This is the final report of the multi-year research project, entitled "Advancement of Latent Trait Theory", which was sponsored by the Office of Naval Research in 1981 through 1987 (N00014-81-C-0569). The first half of the research project was conducted under the above title in June, 1981 through June, 1984. Right after this, in June, 1984, through, Decembei, 1987, the second half of the research continued under the title "Advancement of Latent Trait Theory II". Since the objectives of the two proposed research projects are parallel, i.e, those of the second half are to pursue the objectives of the first half further and in more detail, the present final report will treat them as one long research project, and the accomplishments of these two projects will be integrated and systemized together. These accomplisments include those which have already been published as ONR research reports as well as those still in progress, which will be published in later years as part of more comprehensive research results.

The rest of this chapter will describe papers published or presented during the research period, and related events. The contents of the research accomplishments will be summarized and systematized, and will be described in the succeeding chapters.

[I.1] Research Reports

The following are the ONR research reports that have been published in the present research project.

(1) Information loss caused by noise in models for dichotomous items. *Office of Naval Research Report* 82-1, 1982.

(2) Effect of Noise in the Three-Parameter Logistic Model. *Office of Naval Research Report* 82-2, 1982.

(3) A Latent Trait Model for Differential Strategies in Cognitive Processes. *Office of Naval Research Report* 83-1, 1983.

(4) Information functions for the general model developed for differential strategies in cognitive processes. *Office of Naval Research Report* 83-2.

(5) A general model for the homogeneous case of the continuous response. *Office of Naval Research Report* 83-3, 1983.

(6) Plausibility functions of Iowa Vocabulary Test items estimated by the Simple Sum Procedure of the Conditional P.D.F. Approach. *Office of Naval Research Report* 84-1, 1984.

(7) Comparison of the estimated item parameters of Shiba's Word/Phrase Comprehension Tests obtained by LOGIST 5 and those by the tetrachoric method. *Office of Naval Research Report* 84-2, 1984.

(8) Results of item parameter estimation using Logist 5 on simulated data. *Office of Naval Research Report* 84-3, 1984.

(9) Bias function of the maximum likelihood estimate of ability for discrete item responses. *Office of Naval Research Report* 87-1, 1987.

(10) Final Report: Advancement of latent trait theory. *Office of Naval Research Final Report*, 1988.

[I.2] Advanced Seminar on Latent Trait Theory

As was proposed in the present research, advanced Seminar on Latent Trait Theory was held at the Sheraton Gatlinburg Hotel, Gatlinburg, Tennessee, for four days on March 30 through April 2, 1982.

Lectures were given to approximately forty researchers and graduate students from all over the country, with the principal investigator as the sole speaker. The contents of the lectures were taken, mainly, from her work of the past five years on various topics in Latent Trait Theory, including more general topics such as the method of moments as the least squares solution for fitting polynomials, etc. The topics and contents of this Advanced Seminar are given in appendix A.

The computer package programs were also introduced to the participants of the seminar. These seven package programs are as the following.

(1) TAU TRANSFORMATION: The process of transforming the original latent trait θ to τ , which provides the Old Test with a constant amount of test information. Old Test is a set of test items whose operating characteristics are known and in this case they follow the normal ogive model. The operating characteristics of "unknown" test items are to be estimated, depending partially upon the information provided by the Old Test.

(2) MLE THETA: The process of obtaining the maximum likelihood estimate of θ for each individual examinee from his or her response pattern on the Old Test.

(3) CONDITIONAL MOMENTS MLE: The process of estimating the conditional moments of τ , given its maximum likelihood estimate $\hat{\tau}$, which is transformed from $\hat{\theta}$, i.e., the maximum likelihood estimate of θ . It also includes in the process the approximation of the density function of $\hat{\tau}$ using the method of moments to fit a polynomial to the set of observations.

(4) SIMPLE/WEIGHTED SUM NT: Simple Sum Procedure and Weighted Sum Procedure of the Conditional P.D.F. Approach combined either with the Normal Approach Method or with the Two-Parameter Beta Method, to produce the estimated operating characteristics of the discrete item responses of the "unknown" test items.

(5) PROPORTIONED SUM NT: Proportioned Sum Procedure of the Conditional P.D.F. Approach combined either with the Normal Approach Method or with the Two-Parameter Beta Method, to produce the estimated operating characteristics of the discrete item responses of the "unknown" test items.

(6) CONDITIONAL MOMENTS SUBGROUP: The process of estimating the conditional moments of τ , given its maximum likelihood estimate $\hat{\tau}$, which is transformed from $\hat{\theta}$, i.e., the maximum likelihood estimate of θ , for each discrete item response subgroup of each "unknown" test item. It also includes the approximation of the density function of $\hat{\tau}$ for each subgroup using the method of moments to fit a polynomial to the set of observations.

(7) BIVARIATE P.D.F. NT: Bivariate P.D.F. Approach combined either with the Normal Approach Method or with the Two-Parameter Beta Method, to produce the operating characteristics of the discrete item responses of the "unknown" test items.

Judging from the participants' reactions during and after the Seminar, it is believed that the Seminar gave them a good grasp of the new theoretical and methodological developments made by the principal investigator, i.e., an accomplishment toward the goal of the present research, Advancement of Latent Trait Theory.

[I.3] Invited Conference Addresses

During this period, there were two invited conference paper presentations and one special lecture introducing some of the accomplishments of the principal investigator's research. They are as follows:

(1) Some methods and approaches of estimating the operating characteristics of discrete item responses. *Dr. Frederic M. Lord's Festchrift Conference to Celebrate His Seventieth Birthday*, Educational Testing Service, 1982, Princeton, New Jersey, U. S. A.

(2) Development and application of methods for estimating operating characteristics of discrete item responses without assuming any mathematical form. *1982 Item Response Theory and Computerized Adaptive Conference*, University of Minnesota, 1982, Minneapolis, Minnesota, U. S. A.

(3) Overview of latent trait models. *1987 Annual Meeting of Behaviormetric Society of Japan*, Kyushu University, 1987, Fukuoka, Japan.

At the same conference where the principal investigator presented the paper described in (2), she also served as the discussant to Dr. Roderick P. McDonald's paper, "Unidimensional and multidimensional models for item response theory."

The address described as (3) in the above list was a one hour special lecture overviewing latent trait models. There were more than two hundred Japanese researchers in behaviormetrics among the audience. The summary of the paper is given as Appendix B of this report.

[I.4] Paper Presentations at National and International Conferences

In addition to the three invited papers which were listed in the preceding section, there were other paper presentations at national and international conferences introducing the principal investigator's work. They include ONR contractors' meetings, and are listed below.

(1) *Model Validation, Estimation of Plausibility Functions, Models for Cognitive Processes, and Effect of Noise in the Three-Parameter Logistic Model.* ONR Conference on Model-Based Psychological Measurement, 1983, University of Illinois, Champaign, Illinois, U. S. A.

(2) *A Latent Trait Model for Differential Strategies.* ONR Conference on Action, Attention and Individual Differences in Information Processing, 1984, Haskins Laboratories, New Haven, Connecticut, U. S. A.

(3) *Specification of the Information Provided by Distractors of the Multiple-Choice Test Item and Efficient Ability Estimation.* American Educational Research Association Meeting, New Orleans, 1984. U. S. A.

(4) *Efficient Use of Distractors in Ability Estimation with the Multiple-Choice Test Item.* American Educational Research Association Meeting, New Orleans, 1984. (Coauthorship with Paul S. Changas) U. S. A.

(5) *An Application of Latent Trait Theory in Analyzing the Field Test Results of a Mathematics Proficiency Test.* American Educational Research Association Meeting, New Orleans, 1984. (Coauthorship with Megumi Asako and Allen Knight) U. S. A.

(6) *Further Investigation of the Estimation of the Item Characteristic Function and the Plausibility Functions of the Multiple-Choice Test Item.* ONR Conference on Model Based Measurement, 1984, Educational Testing Service, Princeton, New Jersey, U. S. A.

(7) *A Latent Trait Model When the Item Score Distribution Is Partly Continuous and Partly Discrete.* ONR Conference on Model-Based Psychological Measurement, 1984, Educational Testing Service, Princeton, New Jersey, U. S. A.

(8) *Latent Trait Models Dealing with Continuous Data.* American Educational Research Association Meeting, Chicago, 1985. U. S. A.

(9) *A Content-Based Investigation of Informative Distractors for Multiple-Choice Items of the Iowa Tests of Basic Skills.* American Educational Research Association Meeting, Chicago, 1985. U. S. A. (Coauthorship with Paul S. Changas)

(10) *Expansion of the General Model for the Homogeneous Case of the Continuous Response Level with a Partly Continuous and Partly Discrete Item Score Distribution in the Framework of Latent Trait Theory.* Psychometric Society 50th Anniversary Meeting, 1985, Vanderbilt University, Nashville, Tennessee, U. S. A.

(11) *Latent trait theory as applications of stochastic processes.* Fifteenth Conference on Stochastic Processes and Their Applications, 1985, under the auspices of the Committee for Conferences on Stochastic Processes of the Bernoulli Society. Nagoya University, Nagoya, Japan.

(12) *Effect of the guessing parameter on the estimation of the item discrimination and difficulty parameters when three-parameter logistic model is assumed.* American Educational Research Association Meeting, San Francisco, California, 1986, U. S. A.

(13) *Content-based observation of informative distractors, bias function of the maximum likelihood estimate of the latent trait when item responses are discrete, etc.* ONR Conference on Model-Based Measurement, Gatlinburg, Tennessee, 1986, U. S. A.

(14) *Bias function of the maximum likelihood estimate of the latent trait when item responses are discrete.* American Educational Research Association Meeting, Washington, D. C., 1987, U. S. A.

(15) *Striving for the refinement of the conditional P.D.F. approach for estimating the operating characteristics of discrete responses.* ONR Conference on Model-Based Measurement, Columbia, South Carolina, 1987, U. S. A.

(16) *A robust method of on-line calibration.* American Educational Research Association Meeting, New Orleans, 1988, U. S. A. (Proposed and accepted.)

Out of these paper presentations, the one listed as (11) was made at the international conference held in Nagoya, Japan. Approximately three hundred and sixty researchers, the majority of whom are mathematicians, participated from twenty-five different countries. The principal investigator also chaired one of the sessions by the request of the conference organizer, Professor Takeyuki Hida of Nagoya University.

[I.5] Book Chapters

Some of the principal investigator's works were published as book chapters in two books during this period. They are as follows:

(1) Some methods and approaches of estimating the operating characteristics of discrete item responses. In H. Wainer and S. Messick (Ed.), *Principals of Modern Psychological Measurement: A Festschrift for Frederic M. Lord*, pages 159-182. New Jersey: Lawrence Erlbaum, 1983. New York: Academic Press, 1983.

(2) The constant information model on the dichotomous response level. In David J. Weiss (Ed.) *New Horizons in Testing*, pages 287-308.

[I.6] Other Events

The principal investigator hosted an annual ONR Conference on Model- Based Measurement in 1986, on April 27 through 30, at Park Vista Hotel, Gatlinburg, Tennessee. Approximately forty researchers participated in the conference.

She also gave a seminar on "the On-Line Item Calibration Using Nonparametric Approaches," in July, 1987, at Educational Testing Service.

II Theory and Methods for Estimating the Operating Characteristics of Discrete Item Responses

This part of research aimed at further developments and modifications of the methods and approaches developed by the principal investigator. During the previous four years, 1977 through 1981, the principal investigator had been engaged in the research sponsored by the Office of Naval Research, under the title, "Efficient Methods of Estimating the Operating Characteristics of Item Response Categories and Challenge to a New Model for the Multiple-Choice Item." One of the main outcomes of the research was various methods and approaches for the efficient estimation of the operating characteristics of discrete item responses. They are listed in the summary of the principal investigator's special lecture given at Fukuoka, Japan (Appendix B, page 4), and the computer package programs for these methods are described in [I.2].

Two important features of the principal investigator's approach are the following.

(1) It does not assume any specific mathematical form for the operating characteristic, or the conditional probability, given the latent trait, with which the individual subject gives a specific discrete response.

(2) It does not require a large sample size of individual subjects, i.e., no more than several thousand and sometimes even down to several hundred.

In the present research, computer programs written in the previous research were tested with empirical and simulated data, revised and improved, used for empirical data (cf. Chapter VII), revised again, and so forth. Many variations of these programs were produced. Among others, a series of variations for computerized adaptive testing and on-line item calibration was written. A new method called Lognormal Approach Method was proposed and tested. The bias function of the maximum likelihood estimate was conceived and proposed (cf. Chapter III) partly from the necessity for increasing the accuracy in the operating characteristic estimation, as well as for the general purpose of the advancement of latent trait theory.

[II.1] Conditional P.D.F. Approach Combined with the Normal Approach Method

Out of these different approaches, Bivariate P.D.F. Approach may be the most orthodox one. It has its disadvantages in comparison with the conditional P.D.F. Approach, however, in the sense that: 1) it requires a larger sample size of individual subjects, and 2) its CPU time is substantially greater because of the fact that the estimation has to be done for one item at a time. On the other hand, in spite of the additional approximation involved in the Conditional P.D.F. Approach, in the present research the results of this approach proved to be quite accurate in many cases.

Let θ be the latent trait, or "ability", which assumes any real number. Let τ be the transformed ability which is strictly increasing in θ. In the Conditional P.D.F. Approach, the conditional density, $\phi(\tau \mid \tau_s)$, of τ, given its maximum likelihood estimate $\hat{\tau}_s$ of the subject s, is approximated by some specified probability density function. In so doing, first we estimate the conditional moments of τ, given τ_s, and then use the method of moments for fitting a specific probability density function. Three methods, i.e., Pearson System Method, Two-Parameter Beta Method and Normal Approach Method, have been proposed and tested. Out of these three methods, Pearson System Method provides us with more varieties of "shapes" for the conditional density, including asymmetric ones. Thus it is theoretically more adequate than the other two, i.e., Normal Approach Method and Two-Parameter Beta Method. These two methods have their own advantage, however, for they avoid the use of the third and fourth moments of the conditional distribution of the transformed ability, which are less accurately estimated than the first and second moments.

In our investigation, Normal Approach Method combined either with the Simple Sum Procedure of the Conditional P.D.F. Approach or with the Bivariate P.D.F. Approach was mainly used in the estimation of plausibility functions and in the model validation. The results suggest that, with a sample size as small as several hundred, while there are no visible differences between the two in estimating the item characteristic functions of dichotomous test items, the first combination seems to be better than the second for the estimation of the plausibility functions of distractors and of the operating characteristics of graded item responses. The general success in the use of the Normal Approach Method depends upon the fact that the conditional distribution is indeed approximately normal if the transformed ability distribution is normal, and is approximately truncated normal if the distribution is rectangular, with the truncation negligibly small for the wide range of the maximum likelihood estimate of the transformed ability.

For the reasons described above, after these investigations, Conditional P.D.F. Approach combined with the Normal Approach Method was most frequently used in the present research. Among others, application of this combination to the estimation of the plausibility functions of the wrong alternative answers of the Iowa Tests of Basic Skills items turned out to be very successtul. It will be introduced in Chapter VII.

[II.2] Lognormal Approach Method

As was pointed out in the preceding section, the estimation of the conditional moment of ability, given its maximum likelihood estimate, becomes less accurate as the degree of the moment advances. Thus Pearson System Method must use fairly inaccurately estimated fourth conditional moments, in addition to the better estimated first through third moments. On the other hand, although Normal Approach Methods has an advantage of solely using fairly accurately estimated first and second moments, it has a disadvantage of forcing the estimated conditional density function into symmetry. This forced symmetry could be inappropriate when the population ability distribution departs from normality.

For this reason, it will be a logical direction of research to pursue another method which uses the first, second and third conditional moments, allowing asymmetry to the conditional density function of ability without using the fourth moment. Thus Lognormal Approach Method was developed and proposed. Although it has not been published in a research report yet, it was introduced at the 1987 ONR Conference on Model-Based Measurement (cf. [I.4.15]).

Let μ'_r denote the r-th conditional moment of τ about the origin, given $\hat{\tau}_s$, and μ_r be that of τ about the mean, respectively, i.e.

$$(2.1) \qquad \mu'_r = E(\tau_r \mid \hat{\tau}_s)$$

and

$$(2.2) \qquad \mu_r = E[(\tau - \mu'_1)^r \mid \hat{\tau}_s] \ .$$

In the Lognormal Approach Method, we need the estimates of μ'_1 , μ_2 and μ_3 for each $\hat{\tau}_s$. Let δ denote the skewness index such that

(2.3) $$\delta = \mu_3\, \mu_2^{-3/2} \ .$$

If $\delta \doteq 0$, then the conditional density $\phi(\tau \mid \hat{\tau}_s)$ is approximately symmetric, and it will be approximated by the normal density function with the estimates $\hat{\mu}_1'$ and $\hat{\mu}_2^{1/2}$ as its two parameters, just as we do in the Normal Approach Method.

When $\delta > 0$, the estimated conditional density, $\hat{\phi}(\tau \mid \hat{\tau}_s)$, is given by

(2.4) $$\hat{\phi}(\tau \mid \hat{\tau}_s) \equiv (\tau - \hat{\alpha})^{-1}(2\pi)^{-1/2}\hat{\gamma}^{-1}\, exp[-\{\log(\tau - \hat{\alpha}) - \hat{\beta}\}^2/2\hat{\gamma}^2] \ ,$$

where the estimated three parameters, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$, are obtained through the relationships,

(2.5) $$\begin{aligned} \mu_1' &= \int_\alpha^\infty \tau\phi(\tau \mid \hat{\tau}_s)\, d\tau \\ &= \int_{-\infty}^\infty (e^{\gamma y+\beta} + \alpha)\phi^*(y)\frac{d\tau}{dy}\, dy \\ &= exp\{\beta + (1/2)\gamma^2\} + \alpha \quad , \end{aligned}$$

(2.6) $$\begin{aligned} \mu_2 &= \int_\alpha^\infty (\tau - \mu_1')^2\phi(\tau \mid \hat{\tau}_s)\, d\tau \\ &= \int_\alpha^\infty (\tau - \alpha)^2\phi(\tau \mid \hat{\tau}_s)\, d\tau - (\alpha - \mu_1')^2 \\ &= exp\{2\beta + 2\gamma^2\} - exp\{2\beta + \gamma^2\} \\ &= \nu^2\omega(\omega - 1) \end{aligned}$$

and

(2.7) $$\begin{aligned} \mu_3 &= \int_\alpha^\infty (\tau - \mu_1')^3\phi(\tau \mid \hat{\tau}_s)\, d\tau \\ &= \int_\alpha^\infty (\tau - \alpha)^3\phi(\tau \mid \hat{\tau}_s)\, d\tau + 3\int_\alpha^\infty (\tau - \alpha)^2(\alpha - \mu_1')\phi(\tau \mid \hat{\tau}_s)\, d\tau \end{aligned}$$

$$+ 3\int_{\alpha}^{\infty} (\tau - \alpha)(\alpha - \mu_1')^2 \phi(\tau \mid \hat{\tau}_s)\, d\tau + \int_{\alpha}^{\infty} (\alpha - \mu_1')^3 \phi(\tau \mid \hat{\tau}_s)\, d\tau$$

$$= exp\{3\beta + (9/2)\gamma^2\} - 3\, exp\{3\beta + (5/2)\gamma^2\} + 2\, exp\{3\beta + (3/2)\gamma^2\}$$

$$= \nu^3 \omega^{3/2} (\omega - 1)^2 (\omega + 2) \quad ,$$

where

$$\omega \equiv \{\gamma^2\} \tag{2.8}$$

and

$$\nu \equiv exp(\beta) \quad . \tag{2.9}$$

A similar but somewhat different procedure is taken when $\delta < 0$. In this case, the approximated conditional density function is provided by

$$\hat{\phi}(\tau \mid \hat{\tau}_s) = (\alpha - \tau)^{-1} (2\pi)^{-1/2} \hat{\gamma}^{-1}\, exp[-\{\log(\hat{\alpha} - \hat{\tau}) - \hat{\beta}\}^2 / 2\hat{\gamma}^2] \tag{2.10}$$

where the estimation of the three parameters is conducted through the relationships

$$\begin{aligned} \mu_1' &= \int_{-\infty}^{\alpha} \tau \phi(\tau \mid \hat{\tau}_s)\, d\tau \\ &= \int_{-\infty}^{\infty} \{-e^{(-\gamma y + \beta)}\} \phi^*(y) \frac{d\tau}{dy} dy \\ &= -exp\{\beta + (1/2)\gamma^2\} + \alpha \quad , \end{aligned} \tag{2.11}$$

$$\begin{aligned} \mu_2 &= \int_{-\infty}^{\alpha} (\tau - \mu_1')^2 \phi(\tau \mid \hat{\tau}_s)\, d\tau \\ &= \int_{-\infty}^{\alpha} (\tau - \alpha)^2 \phi(\tau \mid \hat{\tau}_s)\, d\tau - (\alpha - \mu_1')^2 \\ &= exp\{2\beta + 2\gamma^2\} - exp\{2\beta + \gamma^2\} \\ &= \nu^2 \omega (\omega - 1) \end{aligned} \tag{2.12}$$

and

$$(2.13) \qquad \mu_3 = \int_{-\infty}^{\alpha} (\tau - \mu_1')^3 \phi(\tau \mid \hat{\tau}_s)\, d\tau$$

$$= \int_{-\infty}^{\alpha} (\tau - \alpha)^3 \phi(\tau \mid \hat{\tau}_s)\, d\tau + 3\int_{-\infty}^{\alpha} (\tau - \alpha)^2 (\alpha - \mu_1') \phi(\tau \mid \hat{\tau}_s)\, d\tau$$

$$+ 3\int_{-\infty}^{\alpha} (\tau - \alpha)(\alpha - \mu_1')^2 \phi(\tau \mid \hat{\tau}_s)\, d\tau + \int_{-\infty}^{\alpha} (\alpha - \mu_1')^3 \phi(\tau \mid \hat{\tau}_s)\, d\tau$$

$$= -exp\{3\beta + (9/2)\gamma^2\} + 3\, exp\{3\beta + (5/2)\gamma^2\} - 2\, exp\{3\beta + (3/2)\gamma^2\}$$

$$= -\nu^3 \omega^{3/2} (\omega - 1)^2 (\omega + 2) \quad .$$

The actual procedure starts from obtaining the estimate of δ from (2.3), and then that of w through the relationship

$$(2.14) \qquad \delta^2 = (\omega - 1)(\omega + 2)^2 \quad .$$

Then we proceed to estimate the two parameters γ and β concurrently through

$$(2.15) \qquad \gamma = [\log \omega]^{1/2} \quad ,$$

$$(2.16) \qquad \nu = \mu_2^{1/2} \omega^{-1/2} (\omega - 1)^{-1/2}$$

and

$$(2.17) \qquad \beta = \log \nu \quad ,$$

and finally we obtain the estimate of α through

$$(2.18) \qquad \alpha \begin{cases} = \mu_1' + exp\{\beta + (1/2)\gamma^2\} & \delta < 0 \\ = \mu_1' - exp\{\beta + (1/2)\gamma^2\} & \delta > 0 \quad . \end{cases}$$

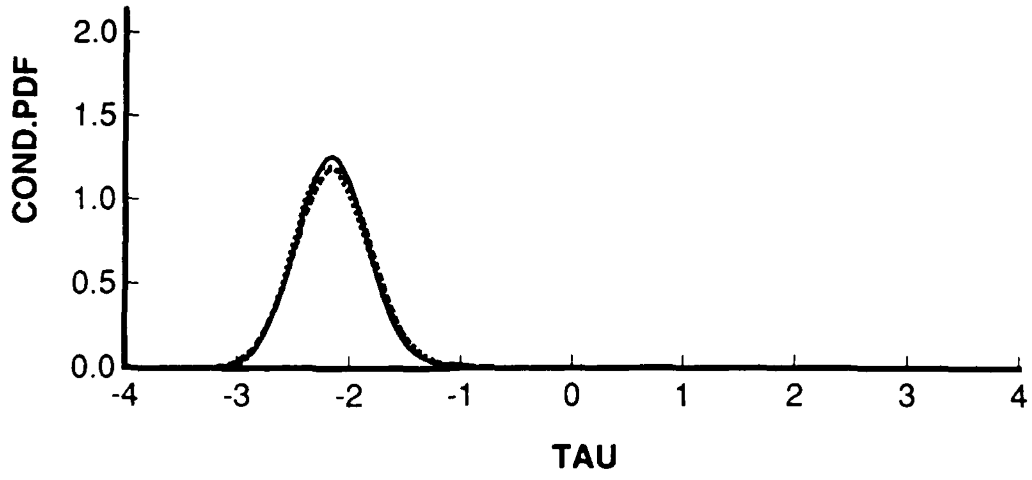
We tested this method implemented in the Conditional P.D.F. Approach with some simulated data, and the results turned out to be at least as good as those obtained by the Normal Approach Method. Figure 2-1 illustrates six examples of the estimated conditional density of τ, given $\hat{\tau}_s$, in comparision with the true density function and the one estimated by the normal approach method. We can see that the improvement is substantial when the true curve is either negatively or positively skewed. This is happening when $\hat{\tau}_s$ is much greater or much less than the mean of $\hat{\tau}_s$.

True appreciation of the method will be reached, however, when it is tested against data having unconditional distributions of τ which are substantially deviated from normality and from uniformity. This will be done in a separate research in the near future.

[II.3] Discussion

This part of research included a substantial amount of computer programming not only for making Lognormal Approach Method accessible but also for modifying and improving the already written package programs. Another orientation was taken to adjust these methods and approaches to the computerized adaptive testing. This is still in the progress, and will be reported in a separate research in the future.

There are many other developments and findings which are not given here. The reader who is interested is directed to the separate research reports and/or to personal conversations with the principal investigator.



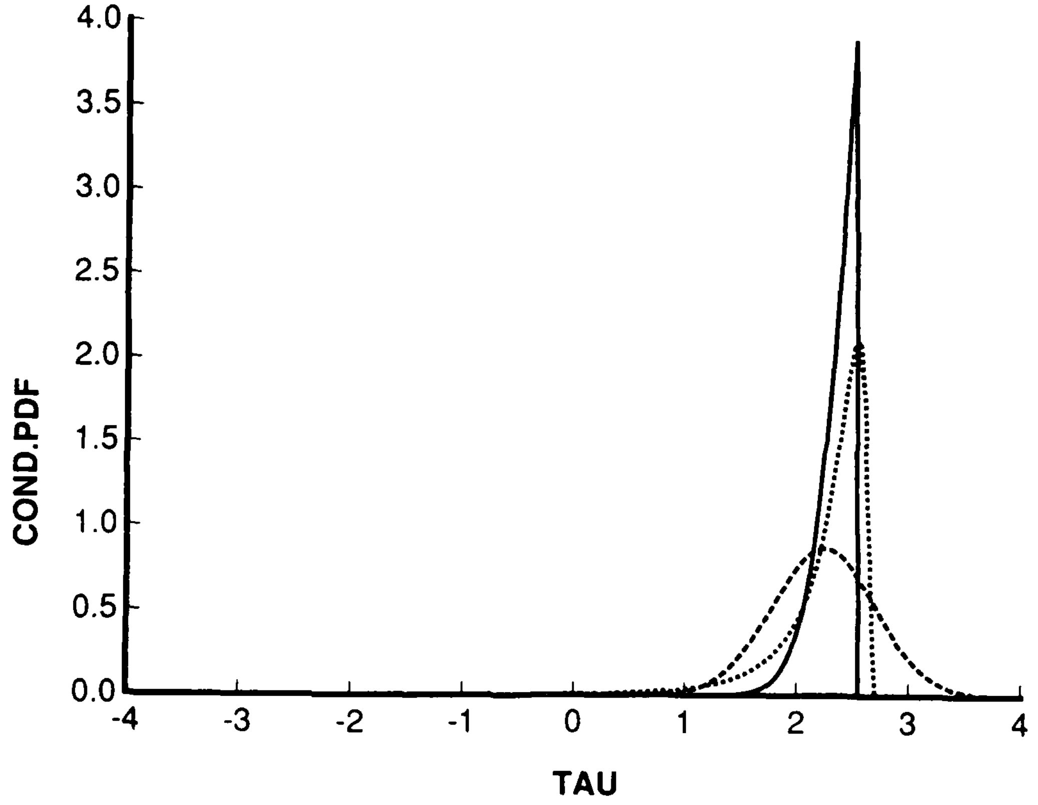


FIGURE 2-1

Six Examples of the Lognormal Curve (Dotted Line) Approximating the Truth Curve (Solid Line) in Comparison with the Normal Density Curve (Dashed Line).

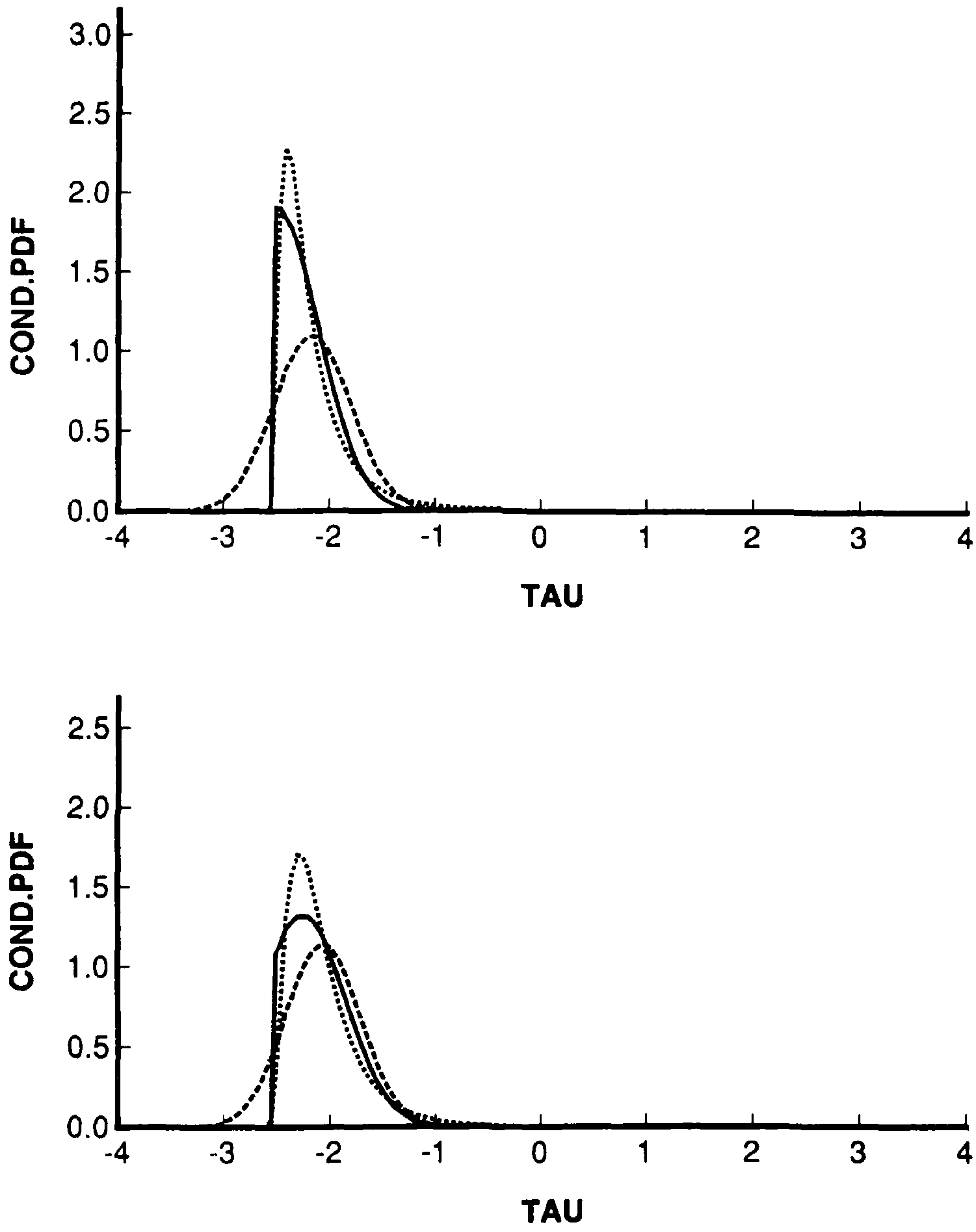


FIGURE 2-1 (Continued)

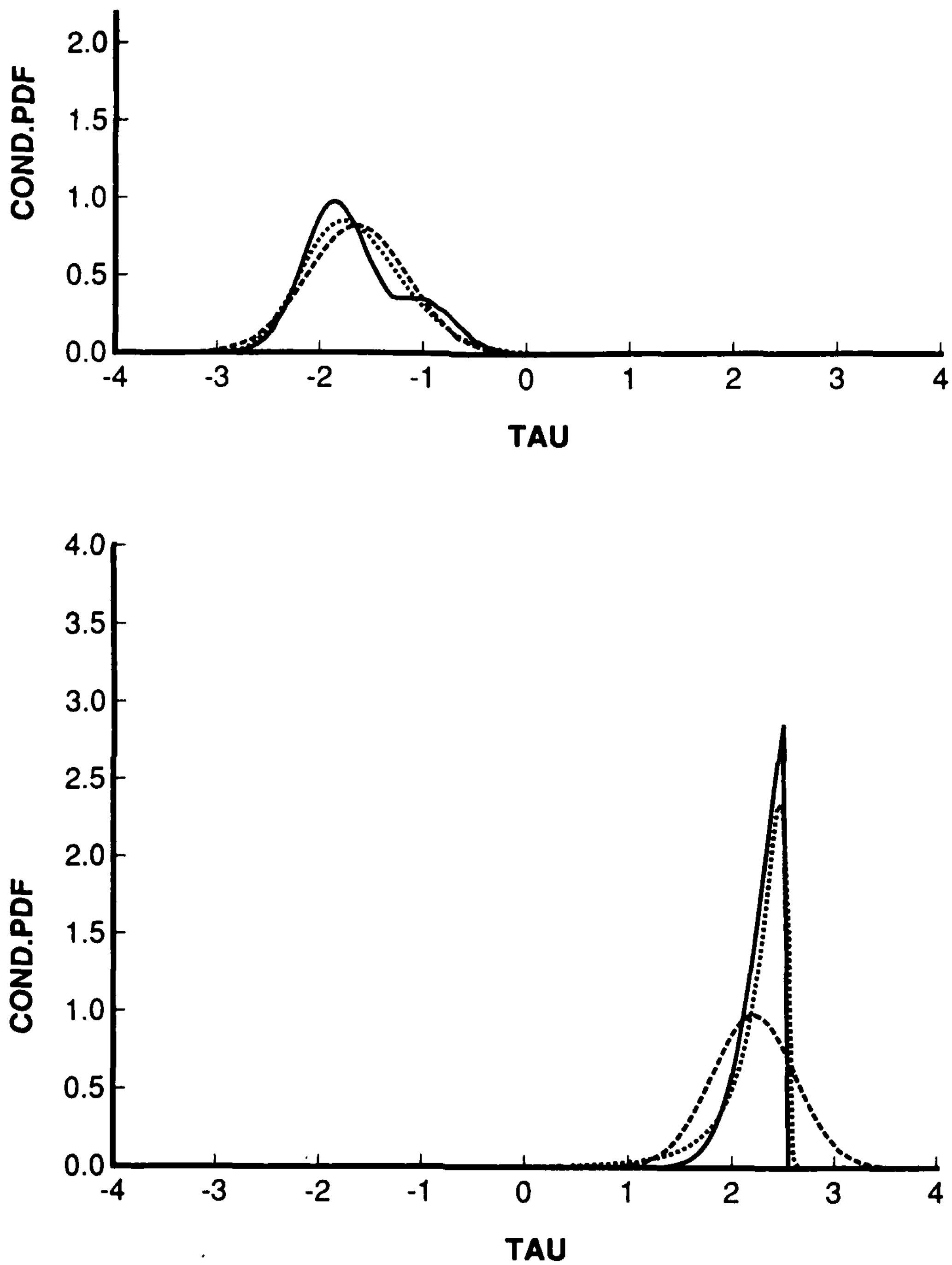


FIGURE 2-1 (Continued)

III Bias Function of the Maximum Likelihood Estimate of Ability in the General Discrete Response Case

In those theory and methods developed for estimating the operating characteristics of discrete item responses, which were described in the preceding chapter, the maximum likelihood estimate $\hat{\theta}$ of ability θ , and also $\hat{\tau}$ of the transformed ability τ play important roles. Since in these methods the asymptotic unbiasedness and normality of the conditional distribution of the maximum likelihood estimate, given the true parameter, is used as approximation, it is of our serious concern whether indeed the maximum likelihood estimate is practically conditionally unbiased or not with actual data for the interval of ability of interest. For this reason and for many others, in the second half of the research period, the bias function of the maximum likelihood estimate of ability was investigated in the general case where item responses are discrete. In this chapter, the outline of its main outcomes will be described. For details and more information, see [I.1.9].

[III.1] Background

Lord has proposed and discussed a bias function of the maximum likelihood estimate in the context of the three-parameter logistic model (cf. Lord, 1983). In so doing, he used Taylor's expansion of the likelihood equation and proceeded from there, obtained an equation which includes the conditional expectation of the discrepancy between the maximum likelihood estimate and the true ability, and ignored all terms of orders higher than n^{-1} , where n indicated the number of items. Let $P_g(\theta)$ be the item characteristic function or item response function in the three-parameter logistic model, which is given by,

$$P_g(\theta) = c_g + (1-c_g)[1+exp\{-Da_g(\theta-b_g)\}]^{-1} \quad , \tag{3.1}$$

where a_g , b_g , and c_g are the item discrimination, difficulty, and guessing parameters, and D is a scaling factor, which is set equal to 1.7 when the logistic model is used as a substitute for the normal ogive model. Lord's bias function $B(\theta)$ can be written as

$$B(\theta) = D[I(\theta)]^{-2}\sum_{g=1}^{n} a_g I_g(\theta)[\Psi_g(\theta)-\frac{1}{2}] \quad , \tag{3.2}$$

where

$$\Psi_g(\theta) = [1+exp\{-Da_g(\theta-b_g)\}]^{-1} \quad , \tag{3.3}$$

and $I_g(\theta)$ and $I(\theta)$ are the item information function and the test information function, respectively, which are given by

$$I_g(\theta) = [P_g'(\theta)]^2[P_g(\theta)\{1-P_g(\theta)\}]^{-1} \quad , \tag{3.4}$$

and

$$I(\theta) = \sum_{g=1}^{n} I_g(\theta) \quad , \tag{3.5}$$

with $P_g'(\theta)$ indicating the first derivative of $P_g(\theta)$ with respect to θ . The former of these two formulae can be given as a special case of the item information function (Samejima, 1969, 1972), which is defined for the general case of discrete responses. (Incidentally, in Lord's paper, $B_1(\hat{\theta})$ is used for this bias function. This is not appropriate, however, since it is a function of θ itself, not of its maximum likelihood estimate $\hat{\theta}$.)

[III.2] Rationale

A similar logic can be adopted for the general case, in which item responses are simply discrete. We assume that there are a finite or an enumerable number of discrete responses k_g 's as possible responses to item g . Thus for the set of n items, we can write for the response pattern V

$$V' = (k_1, k_2, \ldots, k_g, \ldots, k_n) \quad . \tag{3.6}$$

We assume that the operating characteristic $P_{k_g}(\theta)$ is three-times differentiable with respect to θ . By virtue of local independence, we can write for the likelihood function

$$L_V(\theta) = P_V(\theta) = \prod_{k_g \in V} P_{k_g}(\theta) \quad . \tag{3.7}$$

Thus the likelihood equation is given by

$$\frac{\partial}{\partial\theta} \log L_V(\theta) = \sum_{k_g \in V} \frac{\partial}{\partial\theta} \log P_{k_g}(\theta) \equiv 0 \quad . \tag{3.8}$$

We define $\Gamma_{sk_g}(\theta)$ such that

$$\Gamma_{sk_g}(\theta) = \frac{\partial^s}{\partial\theta^s} \log P_{k_g}(\theta) \tag{3.9}$$

for $s = 1, 2, \ldots$ We notice, in particular, that

$$\Gamma_{1k_g}(\theta) = P_{k_g}'(\theta)[P_{k_g}(\theta)]^{-1} = A_{k_g}(\theta) \quad , \tag{3.10}$$

where $A_{k_g}(\theta)$ is the basic function (Samejima, 1969). Let $\Gamma_{sV}(\theta)$ be defined by

$$\Gamma_{sV}(\theta) = \sum_{k_g \in V} \Gamma_{sk_g}(\theta) \tag{3.11}$$

for $s = 1, 2, \ldots$ For a fixed value of θ we can write by Taylor's formula

$$\begin{aligned} \Gamma_{1V}(\hat{\theta}_V) &= \Gamma_{1V}(\theta) + (\hat{\theta}_V - \theta)\Gamma_{2V}(\theta) + (1/2)(\hat{\theta}_V - \theta)^2\Gamma_{3V}(\theta) \\ &\quad + (1/6)(\hat{\theta}_V - \theta)^3\Gamma_{4V}(\theta) + (1/24)(\hat{\theta}_V - \theta)^4\Gamma_{5V}(\xi) = 0 \quad , \end{aligned} \tag{3.12}$$

where ξ is some value between θ and $\hat{\theta}_V$.

Since we have

(3.13)
$$\sum_{k_g} P_{k_g}(\theta) = 1 \ ,$$

we obtain

(3.14)
$$\sum_{k_g} \frac{\partial^t}{\partial \theta^t} P_{k_g}(\theta) = 0$$

for $t = 1, 2, \ldots$ Equation (3.14) will be helpful in following the mathematical derivations which are needed in obtaining the bias function. Let $\gamma_{sg}(\theta)$ be the conditional expectation of $\Gamma_{sk_g}(\theta)$, given θ , which can be written as

(3.15)
$$\gamma_{sg}(\theta) = E[\Gamma_{sk_g}(\theta) \mid \theta] = \sum_{k_g} \Gamma_{sk_g}(\theta)\, P_{k_g}(\theta) \ .$$

In particular, we have from (3.10) and (3.14)

(3.16)
$$\gamma_{1g} = \sum_{k_g} P'_{k_g}(\theta) = 0 \ .$$

We further define $\gamma_s(\theta)$, $\epsilon_{sk_g}(\theta)$ and $\epsilon_{sV}(\theta)$ such that

(3.17)
$$\gamma_s(\theta) = (1/n) \sum_{g=1}^{n} \gamma_{sg}$$

for $s = 1, 2, \ldots$,

(3.18)
$$\epsilon_{sk_g}(\theta) = \Gamma_{sk_g}(\theta) - \gamma_{sg}(\theta)$$

and

(3.19)
$$\epsilon_{sV}(\theta) = (1/n) \sum_{k_g \in V} \epsilon_{sk_g}(\theta) \ ,$$

respectively. For the conditional expectation of $\epsilon_{sV}(\theta)$, given θ , we obtain

(3.20)
$$E[\epsilon_{sV}(\theta) \mid \theta] = \sum_{V} \epsilon_{sV}(\theta)\, P_V(\theta) = \gamma_s(\theta) - \gamma_s(\theta) = 0 \ .$$

With these definitions of $\gamma_s(\theta)$ and $\epsilon_{sV}(\theta)$ and from (3.12) we have

$$\epsilon_{1V}(\theta) + (\hat{\theta}_V - \theta)[\gamma_2(\theta) + \epsilon_{2V}(\theta)] + (1/2)(\hat{\theta}_V - \theta)^2[\gamma_3(\theta) + \epsilon_{3V}(\theta)] + (1/6)(\hat{\theta}_V - \theta)^3[\gamma_4(\theta) + \epsilon_{4V}(\theta)] + (1/24)(\hat{\theta}_V - \theta)^4\Gamma_{5V}(\theta) = 0 \quad , \tag{3.21}$$

and proceeding from here by taking the conditional expectation of each term with respect to V , given θ , and ignoring all terms whose orders are higher than n^{-1} , we obtain

$$E[\epsilon_{1V}(\theta) \mid \theta] + \gamma_2(\theta)E[\hat{\theta}_V - \theta \mid \theta] + E[(\hat{\theta}_V - \theta)\epsilon_{2V}(\theta) \mid \theta] + (1/2)\gamma_3(\theta)E[(\hat{\theta}_V - \theta)^2 \mid \theta] \doteq 0 \quad . \tag{3.22}$$

It is obvious from (3.20) that the first term on the left hand side of (3.22) disappears. As for the fourth and last term in (3.22) we can use the asymtotic variance of the distribution of the maximum likelihood estimate as the approximation to its last factor, i.e.,

$$E[(\hat{\theta}_V - \theta)^2 \mid \theta] \doteq [I(\theta)]^{-1} \quad . \tag{3.23}$$

Thus all that is left to do is to evaluate the third term on the left hand side of (3.22) in the general framework. From this we obtain

$$E[(\hat{\theta}_V - \theta)\epsilon_{2V}(\theta) \mid \theta] = (1/n)[I(\theta)]^{-1}\sum_{g=1}^{n}\sum_{k_g} A_{k_g}(\theta)[P''_{k_g}(\theta) - A_{k_g}(\theta)P'_{k_g}(\theta)] \quad , \tag{3.24}$$

where $P'_{k_g}(\theta)$ and $P''_{k_g}(\theta)$ indicate the first and second derivatives of $P_{k_g}(\theta)$, with respect to θ respectively. Substituting (3.15), (3.17), (3.23) and (3.24) into (3.22) and rearranging, we obtain for the bias function of the maximum likelihood estimate

$$\begin{aligned} B(\theta) = E[\hat{\theta}_V - \theta \mid \theta] &= -(1/2)[I(\theta)]^{-2}\sum_{g=1}^{n}\sum_{k_g} A_{k_g}(\theta)P''_{k_g}(\theta) \\ &= -(1/2)[I(\theta)]^{-2}\sum_{g=1}^{n}\sum_{k_g} P'_{k_g}(\theta)P''_{k_g}(\theta)[P_{k_g}(\theta)]^{-1} \quad . \end{aligned} \tag{3.25}$$

On the graded response level where item score x_g assumes successive integers, 0 through m_g , each k_g in (3.25) must be replaced by x_g . On the dichotomous response level, it can be reduced to the form

$$B(\theta) = E[\hat{\theta}_V - \theta \mid \theta] = (-1/2)[I(\theta)]^{-2}\sum_{g=1}^{n} I_g(\theta)P''_g(\theta)[P'_g(\theta)]^{-1} \quad , \tag{3.26}$$

with $P_g''(\theta)$ indicating the second derivative of $P_g(\theta)$ with respect to θ . This includes Lord's bias function in the three-parameter logistic model as a special case. In the normal ogive model, the item characteristic function is given by

$$P_g(\theta) \; = \; (2\pi)^{-1/2} \int_{-\infty}^{a_g(\theta - b_g)} e^{-u^2/2}\, du \tag{3.27}$$

where a_g and b_g are the item discrimination and difficulty parameters, respectively. From (3.27), we can write for the first and second derivatives of $P_g(\theta)$ with respect to θ

$$P_g'(\theta) \; = \; a_g (2\pi)^{-1/2} e^{-a_g^2(\theta - b_g)^2/2} \quad , \tag{3.28}$$

and

$$P_g''(\theta) \; = \; -a_g^2(\theta - b_g) P_g'(\theta) \quad , \tag{3.29}$$

respectively. Substituting (3.28) and (3.29) into (3.26) and rearranging, we obtain for the bias function

$$B(\theta) \; = \; (1/2)[I(\theta)]^{-2} \sum_{g=1}^{n} a_g^2(\theta - b_g) I_g(\theta) \quad . \tag{3.30}$$

In the (two-parameter) logistic model, the item characteristic function is given by

$$P_g(\theta) \; = \; [1 + e^{-Da_g(\theta - b_g)}]^{-1} \quad , \tag{3.31}$$

which is the same as $\Psi_g(\theta)$ in (3.3). The bias function is the same as (3.2) , therefore, by obtaining $I_g(\theta)$ and $I(\theta)$ by setting $c_g = 0$.

[III.3] Bias Function and Amount of Test Information

We shall introduce some examples now. In developing nonparametric approaches and methods of estimating the operating characteristics, or the conditional probabilities, given ability θ , which are assigned to separate discrete item responses, a set of simulated data has been used for testing these approaches and methods, in which 35 graded test items following the normal ogive model with three item score categories each are hypothesized as the Old Test (cf. Samejima, 1977, 1981). The square root of the test information function of this Old Test is shown as the upper solid curve in Figure 3-1. The bias function, which was computed through (3.25), is also shown in the same figure as the lower solid curve. We can see in this figure that for the interval of θ covering $(-4, 4)$ the bias of the maximum likelihood estimate is practically zero, i.e., the MLE of ability is practically unbiased for this range of θ . Thus, one of the necessary conditions to justify the use of the asymptotic normality as the approximation for the conditional distribution of MLE, given θ , is satisfied.

We notice that for the range of θ , $(-3, 3)$, the square root of the test information function of this Old Test assumes approximately a constant value, 4.65 , and we have already seen that for the wider range of θ the bias function assumes, practically, zero. It is interesting to note that the bias

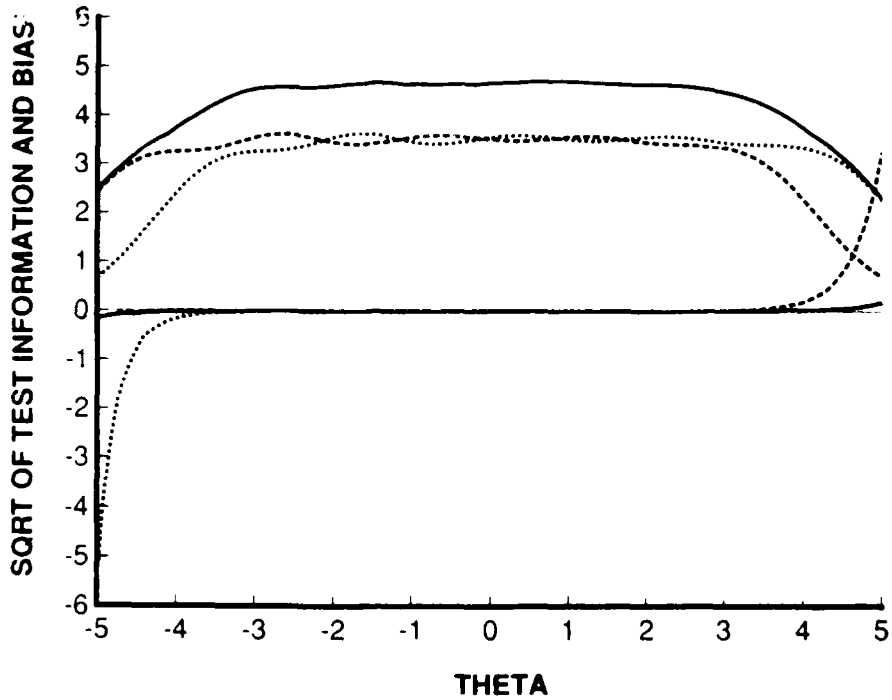


FIGURE 3-1

The Square Root of the Test Information Function and the MLE Bias Function of the Old Test of Thirty-Five Graded Items (Upper and Lower Solid Curves) and Those of the Two Redichotomised Tests (Dashed and Dotted Pairs of Curves, Respectively).

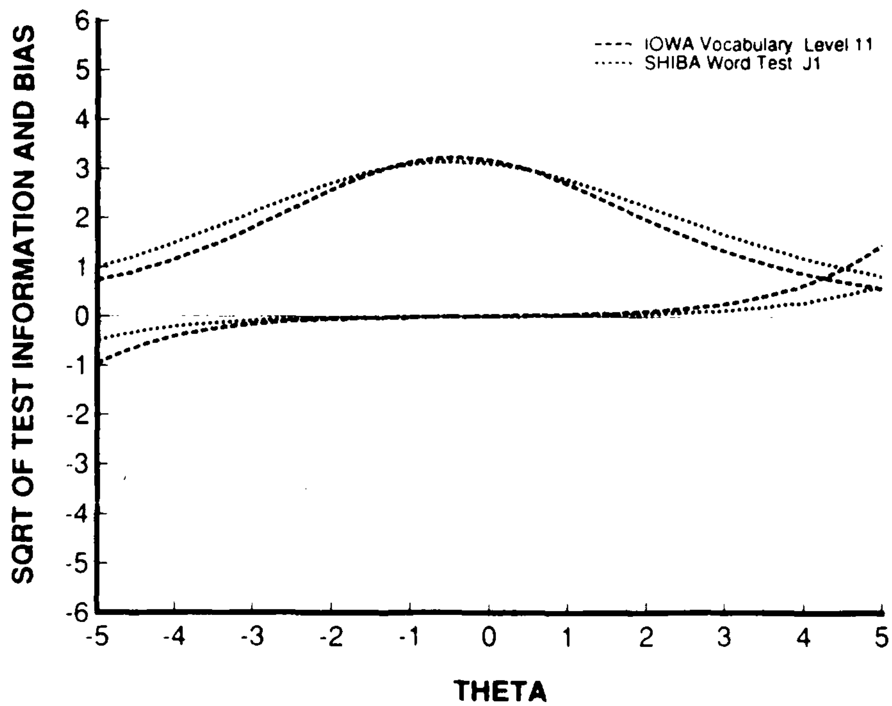


FIGURE 3-2

The Square Root of the Test Information Function and the MLE Bias Function of the Iowa Level 11 Vocabulary Subtest (Upper and Lower Dashed Curves) and Those of Shiba's Word/Phrase Comprehension Test J1 (Upper and Lower Dotted Curves).

starts showing up both positively and negatively when the square root of test information drops lower than a critical value, which is approximately 3.2 , or the test information function drops lower than approximately 10 . In order to pursue this relationship, two more sets of these two functions are also shown by dashed and dotted curves in Figure 3-1. These two sets were created by redichotomizing the graded items of the Old Test, using the first and second sets of the difficulty parameters, respectively. We can see that for the wide range of θ the square root of the test information is substantially less than that of the original Old Test, which is the natural consequence of redichotomizing the items. We notice that for each of these two, the square root of the test information function is barely greater than 3.2 for a wide range of θ , and the bias is practically nil. Again the bias appears both positively and negatively when the square root of the test information function drops lower than approximately 3.2 . If we tolerate the biases of ±0.1 , then the critical value of the square root of test information will approximately be 2.75 , or that of the test information function approximately 7.5 . When the square root of test information drops less than 2.0 , the bias turns out to be substantially large.

Figure 3-2 presents similar results by dashed and dotted curves, respectively, which are based upon two sets of empirical data. The first set is the results of the Level 11 Vocabulary Subtest of 43 items of the Iowa Tests of Basic Skills collected for 2,356 school children of approximately age eleven, and the second set is those of the Test J1 of Shiba's Word/Phrase Comprehension Tests of 55 items collected for 2,259 junior high school students (cf. Samejima, 1981). Both sets of operating characteristics are estimated by assuming the normal ogive model. In these two cases, the critical value of square root of test information when we tolerate biases of ±0.1 turned out to be less, i.e., approximately 1.75 , or the critical value of the test information function is approximately 3.0 . These differences seem to have something to do with the fact that in the Old Test there are only 35 test items with the average discrimination parameter as high as 1.70 , while there are as many as 43 and 55 items in the Iowa Subtest and Shiba's J1 Test with the average values of discrimination parameters 0.601 and 0.538 , respectively for tha Iowa Level 11 Vocabulary subtest.

We can see in this figure that the bias is practically nil for tha range of θ , $(-2,2)$, where approximately 95 percent of the subjects are located. This interval is even wider for Shiba's test J1. This fact proves exellence of the tests in this aspect. These two tests were more thoroughly analysed in the present research project, and these results will be obtained in chapters VII and VIII, respectively.

[III.4] Increment in Bias Caused by Random Guessing

The two graphs in Figure 3-3 show the increment in bias caused by the guessing parameters for the Iowa Level 11 Vocabulary Subtest and Shiba's J1 Test, respectively. In each graph, the solid curve indicates the bias function based upon the logistic model, while the dashed and dotted curves are the bias functions based upon the three-parameter logistic model, with the guessing parameters, 0.20 and 0.25 , respectively, added to the same discrimination and difficulty parameters of each item. It is obvious that a substantial increment in bias is caused by the addition of the guessing parameter, especially on the lower levels of ability.

[III.5] Adaptive Testing

Observations made in the previous sections provide us with ideas how things go in adaptive testing. First of all, in order to reach the practical unbiasedness in estimating the individual subject's ability in adaptive testing, we need to make sure that a sufficient amount of test information has been reached for each individual subject, before terminating the presentation of new items. We can control it easily, if we use the amount of test information as the criterion for the termination of presenting new items, or as the "stopping rule". If the items follow the normal ogive or logistic model in the adaptive testing situation, for subjects of intermediate ability levels it is likely that on the initial stage the item difficulty parameters

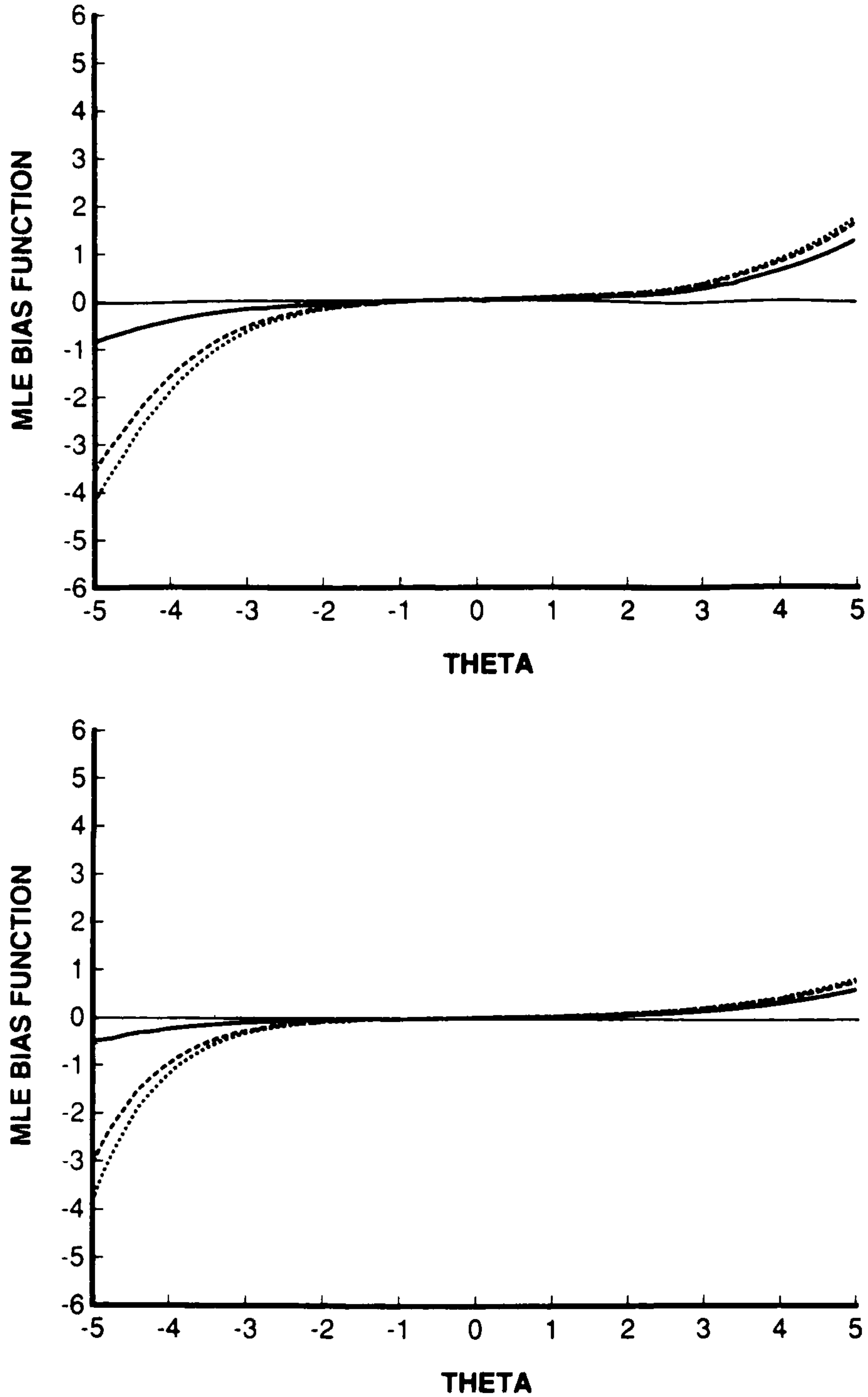


FIGURE 3-3

MLE Bias Function Based upon the Three-Parameter Logistic Model with the Guessing Parameters 0.20 (Dashed Curve) and 0.25 (Dotted Curve), Respectively, in Comparison with the One Based upon the (Two-Parameter) Logistic Model, for the Iowa Level 11 Vocabulary Subtest (Upper Graph) and for Shiba's Word/Phrase Comprehension Test J1 (Lower Graph).

fluctuate both negatively and positively around the subject's true ability level, and consequently, the biases of negative and positive directions are cancelled out, since an item pool usually has plenty of items of intermediate difficulties. In such a case we do not have to worry too much about the influence of the initial items on the eventual bias of the ability estimate. When the maximum likelihood estimate has started being more or less stabilized, chances are slim that the additional item causes a substantial bias, provided that the program is written in such a way that an item of a large amount of information at the current estimated ability level will be presented next, and that the item pool has a sufficient number of items whose difficulty levels are arround the subject's true ability level. There is greater possibility that the examinee obtains a biased ability estimate if his ability level is close to either end of the configuration of difficulty parameters, since biases caused by the initially presented items are not likely to cancel themselves out, and, moreover, there may not be a sufficient number of items whose difficulty levels are close to his ability level.

If the item pool consists of items following the three-parameter normal ogive or logistic model, the effect of random guessing on the amount of bias can be substantial, especially on the lower levels of ability. In such a case, it is imperative to include many easy items in the item pool.

In any case, the bias function can be a good indicator in evaluating the item pool, if we use it wisely and effectively. Those results that were described in previous sections will give us information and suggestions as to how to improve an existing item pool.

[III.6] Discussion

It has been observed that: 1) the amount of bias of the maximum likelihood estimate increases with the decrease of the amount of test information, and there seems to be a relatively simple relationship between the two; 2) on the other hand, it seems that the configuration of the discrimination and difficulty parameters within a test and the *number of items* affect the amount of bias; and 3) random guessing increases the amount of bias especially on the lower levels of ability. A usefulness of the bias function is seen in developing theory and methodologies using the normal approximation of the conditional distribution of the MLE of ability, given θ , as we have seen in the nonparametric estimation of the operating characteristics of discrete item responses.

There are many other developments and observations concerning the MLE bias function that are not presented here. Among others, they include comparison of different models, the effect of the discrimination parameters on the bias function, the effect of the number of items on the bias function, the bias function after the scale was transformed including the general case of discrete responses and the equivalent item case on the dichotomous response level, and the effect of the scale transformation to generate a constant test information.

References

[1] Lord, F. M. Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika,* 48, 1983, 233-245.

[2] Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph,* No. 17, 1969.

[3] Samejima, F. A general model for free-response data. *Psychometrika Monograph,* No. 18, 1972.

[4] Samejima, F. A method of estimating item characteristic functions using the maximum likelihood estimate of ability. *Psychometrika,* 42, 1977, 163-191.

[5] Samejima, F. and Livingston, P.S. Method of moments as the least squares solution for fitting a polynomial. *ONR Research Report,* 79-2, 1979.

[6] Samejima, F. Final Report: Efficient methods of estimating the operating characteristics of item response categories and challenge to a new model for the multiple-choice item. *Final Report of N00014-77-C-0360,* Office of Naval Research, 1981.

IV Constancy in Item Information and the Information Loss Caused by Noise on the Dichotomous Response Level

Researchers tend to consider that dichotomous items with high discrimination parameters are "good" items, and those with low discrimination parameters are "bad" ones. This is not necessarily true, however. If, for example, in our item pool of binary test items most items are of similar levels of difficulty, items with low discrimination parameters will be more informative and useful for testing individual subjects on the levels of latent trait, or "ability", which substantially depart from these levels in either positive or negative direction.

The principal investigator has pointed out (Samejima, 1979) that there is a constancy in the amount of information given by items regardless of the values of their discrimination parameters, provided that these items have the same type of item characteristic functions. This was discussed in the unidimensional latent space. The constancy exists in the square root of the item information function integrated for the entire range of ability θ . Among others, it has been shown that, if the item characteristic function is strictly increasing in θ with zero and unity as its lower and upper asymptote, respectively, as is the case with the normal ogive model, logistic model, linear model, etc., this total area under the square root of the item information function equals π , regardless of mathematical formulae representing particular models. It has also been shown that, if the model has a lower asymptote greater than zero, as is the case with the three-parameter normal ogive and logistic models, etc., the constancy still exists, but the area decreases as the lower asymptote increases. It will be worthwhile to investigate this constancy of item information across different models and item parameters in a more general framework, and also to investigate the amount of information loss caused by noise, such as the guessing parameter in the three-parameter normal ogive or logistic model, etc. One reason for the necessity of such a research is the fact that the three-parameter logistic model has been applied by so many researchers without a deep enough understanding of the model. Another reason is that we need to know more about different types of models in order to use them for different purposes of research. One such example of necessities will be given in modeling differential strategies in cognitive processes, which will be outlined in chapter V.

The principal investigator pursued these topics described above and in this chapter its summary will be presented. For further detail and more information, see [I.1.1] and [I.1.2].

[IV.1] Four Types of Models for Dichotomous Test Items

We assume that the item characteristic function, $P_g(\theta)$, is strictly increasing in ability θ , for the interval

$$\underline{\theta} < \theta < \overline{\theta} \ , \tag{4.1}$$

where $\underline{\theta}$ and $\overline{\theta}$ may be negative and positive infinities, respectively, or finite numbers. This interval, $(\underline{\theta}, \overline{\theta})$, can either be the whole range of ability θ which is common for all items, or a subinterval specified for a particular item. Let c_{g1} and c_{g2} denote the lower and upper asymptotes of the item characteristic function, where

$$0 \le c_{g1} < c_{g2} \le 1 \ . \tag{4.2}$$

Four types of models, Types A, B, C and D, are considered in this research, which are represented by the general formula for the item characteristic function $P_g(\theta)$ such that

$$(4.3) \qquad P_g(\theta) = c_{g1} + (c_{g2} - c_{g1})\Psi_g(\theta) \ ,$$

where $\Psi_g(\theta)$ is a strictly increasing function of θ for the interval $(\underline{\theta}, \overline{\theta})$, with zero and unity as its lower and upper asymptotes, respectively. These four types of models are distinguished from each other by the values of c_{g1} and c_{g2} which are the listed below.

Type A: $0 = c_{g1} < c_{g2} = 1$

Type B: $0 < c_{g1} < c_{g2} = 1$

Type C: $0 = c_{g1} < c_{g2} < 1$

Type D: $0 < c_{g1} < c_{g2} < 1$

Figure 4-1 presents a set of examples of these four types where $\Psi_g(\theta)$ is the item characteristic function of the normal ogive model, which is given by

$$(4.4) \qquad \Psi g(\theta) = (2\pi)^{-1/2} \int_{-\infty}^{a_g(\theta - b_g)} e^{-u^2/2} \, du \ ,$$

with the parameters, $a_g = 1.0$ and $b_g = 0.0$. For Types B and D, we have $c_{g1} = 0.2$, and, for Types C and D, $c_{g2} = 0.8$. We notice that the example for Type A which is given in Figure 4-1 is the normal ogive model itself, and that for Type B is the three-parameter normal ogive model. We have no specific models of Types C and D which are commonly used yet. As will be pointed out in Chapter V, however, models of Type C, in particular, have important roles in dealing with items of multi-correct answers and in modeling differential strategies in cognitive processes. Investigating the characteristics of this type of models will be just as important, therefore, for further advancement of latent trait theory.

[IV.2] Information Loss

Let Q denote the total information, which is defined by

$$(4.5) \qquad Q = \int_{-\infty}^{\infty} [I_g(\theta)]^{1/2} \, d\theta \ ,$$

where $I_g(\theta)$ is the item information function for which we can write

$$(4.6) \qquad I_g(\theta) = [\frac{\partial}{\partial \theta} P_g(\theta)]^2 [P_g(\theta)]^{-1} [1 - P_g(\theta)]^{-1} \ .$$

Following a sequence of logics and mathematics, we obtain for the total information

$$(4.7) \qquad Q = 2[\tan^{-1}\{\frac{c_{g2}}{1 - c_{g2}}\}^{1/2} - \tan^{-1}\{\frac{c_{g1}}{1 - c_{g1}}\}^{1/2}] \ .$$

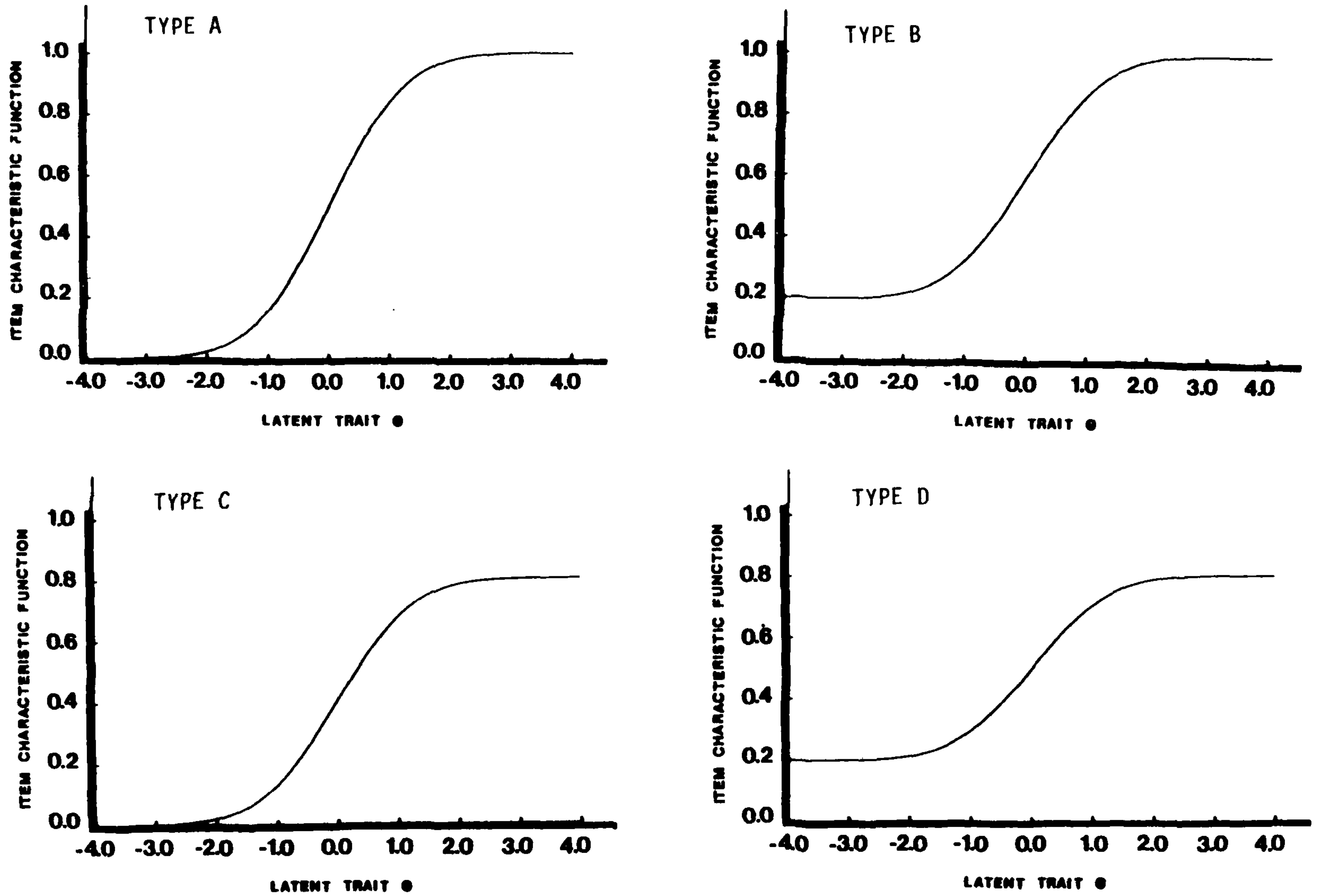


FIGURE 4-1

Examples of Item Characteristic Functions of Types A, B, C and D.

It is obvious from (4.7) that, when item g belongs to Type A, as is the case with the normal ogive model, logistic model, linear model, constant information model, etc., the second term in the second factor of the right hand side of (4.7) disappears, and also the first term takes on the maximal value, $\pi/2$. Thus we have

$$Q = \pi \ , \tag{4.8}$$

the result which is consistent with our previous finding (Samejima, 1979). We can also see from (4.7) that, as c_{g1} departs from zero, and c_{g2} from unity, the total information Q becomes progressively smaller than π. In the three examples shown in Figure 4-1, we obtain $Q = 2.214$ for both Types B and D, and $Q = 1.287$ for Type D. Also we can easily see from (4.7) that the total information Q assumes the same value for Types B and C whenever $c_{g1} = 1 - c_{g2}$.

[IV.3] Basic Functions and Item Response Information Functions

There are certain models for binary items which assure the existence of a unique maximum for the likelihood function of every possible response pattern, such as normal ogive model, logistic model, etc. In fact, except for the two extreme response patterns in which the binary item score u_g assumes zero for all items, and unity for all items, respectively, the likelihood function has a unique local maximum in those models. It has been pointed out (Samejima, 1969, 1972) that a sufficient condition for the unique maximum is: 1) that the basic function, which is given by

$$A_{u_g}(\theta) = (-1)^{u_g+1} \frac{\partial}{\partial\theta} \log P_g(\theta) \qquad u_g = 0, 1 \ . \tag{4.9}$$

is strictly decreasing in θ throughout its whole range, and 2) that its upper asymptote is non-negative and its lower asymptote is non-positive. For brevity, we shall call it the unique maximum condition. This condition implies that the item response information function is positive except, at most, at an enumerable number of points of θ. It has been shown (Samejima, 1972, 1973b) that the three-parameter logistic model does not satisfy the unique maximum condition, and that the likelihood function for certain response patterns has more than one modal point. The same is true with the three-parameter normal ogive model.

There also are models of Type A which do not satisfy the unique maximum condition, which is exemplified by the linear model. For simplicity, let $\Psi_g'(\theta)$ denote the first partial derivative of $\Psi_g(\theta)$ with respect to θ, and $\Psi_g''(\theta)$ be the second partial derivative. We can write for the basic function

$$A_{u_g}(\theta) \begin{cases} = -(c_{g2} - c_{g1})\Psi_g'(\theta)[(1 - c_{g1}) - (c_{g2} - c_{g1})\Psi_g(\theta)]^{-1} & u_g = 0 \\ \\ = (c_{g2} - c_{g1})\Psi'(\theta)[c_{g1} + (c_{g2} - c_{g1})\Psi_g(\theta)]^{-1} & u_g = 1 \ . \end{cases} \tag{4.10}$$

for the general form of the item characteristic function which is specified by (4.3). The item response information function $I_{u_g}(\theta)$ is defined by

$$I_{u_g}(\theta) = -\frac{\partial}{\partial\theta} A_{u_g}(\theta) \ , \tag{4.11}$$

and from (4.3) and (4.10) we can write

$$
(4.12) \qquad I_{u_g}(\theta) \begin{cases} = [(c_{g2} - c_{g1})^2[\{\Psi_g'(\theta)\}^2 + \{1 - \Psi_g(\theta)\}\Psi''(\theta)] + (1 - c_{g2})(c_{g2} - c_{g1}) \\ \qquad \Psi_g''(\theta)][(1 - c_{g1}) - (c_{g2} - c_{g1})\Psi_g(\theta)]^{-2} & u_g = 0 \\ = \{(c_{g2} - c_{g1})^2[\{\Psi_g'(\theta)\}^2 - \Psi_g(\theta)\Psi''(\theta)] - c_{g1}(c_{g2} - c_{g1})\Psi_g''(\theta)\} \\ \qquad [c_{g1} + (c_{g2} - c_{g1})\Psi_g(\theta)]^{-2} & u_g = 1 \ . \end{cases}
$$

These formulae can be simplified for Types A, B and C by substituting c_{g1} by zero and/or c_{g2} by unity.

If we specify $\Psi_g(\theta)$ by the logistic function such that

$$
(4.13) \qquad \Psi_g(\theta) = [1 + e^{\{-Da_g(\theta - b_g)\}}]^{-1} \ ,
$$

then we have

$$
(4.14) \qquad \Psi_g'(\theta) = D\,a_g\,\Psi_g(\theta)[1 - \Psi_g(\theta)] \ ,
$$

and

$$
(4.15) \qquad \Psi_g''(\theta) = D^2\,a_g^2\,\Psi_g(\theta)[1 - \Psi_g(\theta)][1 - 2\,\Psi_g(\theta)] \ .
$$

Figure 4-2 presents four examples of the item response information functions of Types A, B., C and D, with specified item parameters.

As we can see in these examples, in certain cases the item response information function assumes negative values for a specific interval of θ. Let $\underline{\theta}_g$ denote the critical value of θ below which the item response information function of Type B or D assumes negative values for $u_g = 1$, and $\overline{\theta}_g$ be the one above which the item response information function of Type C or D takes on negative values for $u_g = 0$. In general, we can write

$$
(4.16) \qquad \underline{\theta}_g = -(2\,Da_g)^{-1}[\log c_{g2} - \log c_{g1}] + b_g
$$

$$
(4.17) \qquad \overline{\theta}_g = (2\,Da_g)^{-1}[\log(1 - c_{g1}) - \log(1 - c_{g2})] + b_g \ .
$$

These critical values are shown in Figure 4-2.

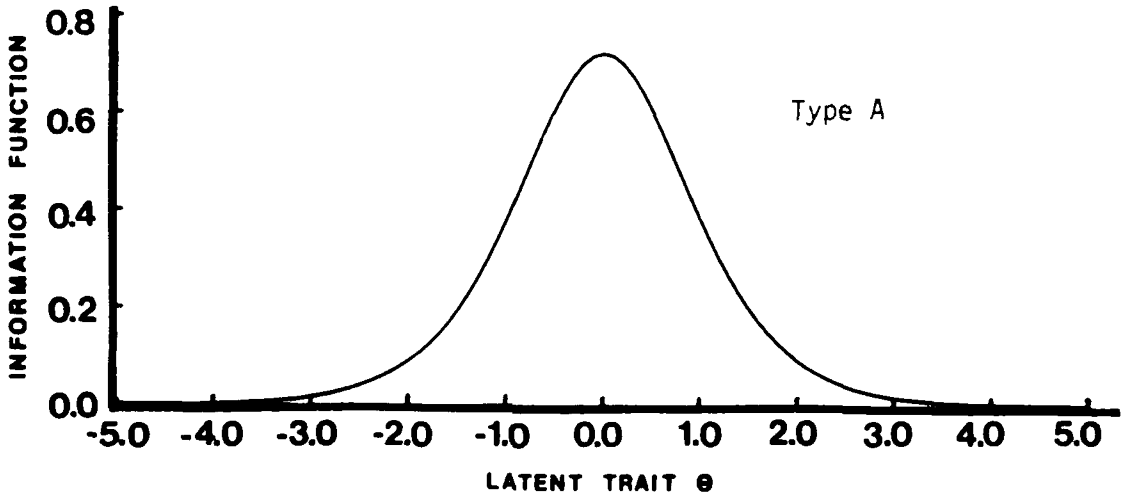


FIGURE 4-2

Two Item Response Information Functions (Solid Lines) and Item Information Function (Dashed Line) of Each of Type A, B, C and D Models: Logistic Function Is Used for $\Psi_g(\theta)$ with $D = 1.7$, $a_g = 1.00$ and $b_g = 0.00$. These Three Curves Overlap for Type A. Values of c_{g1} and/or c_{g2} are specified for Types B, C and D.

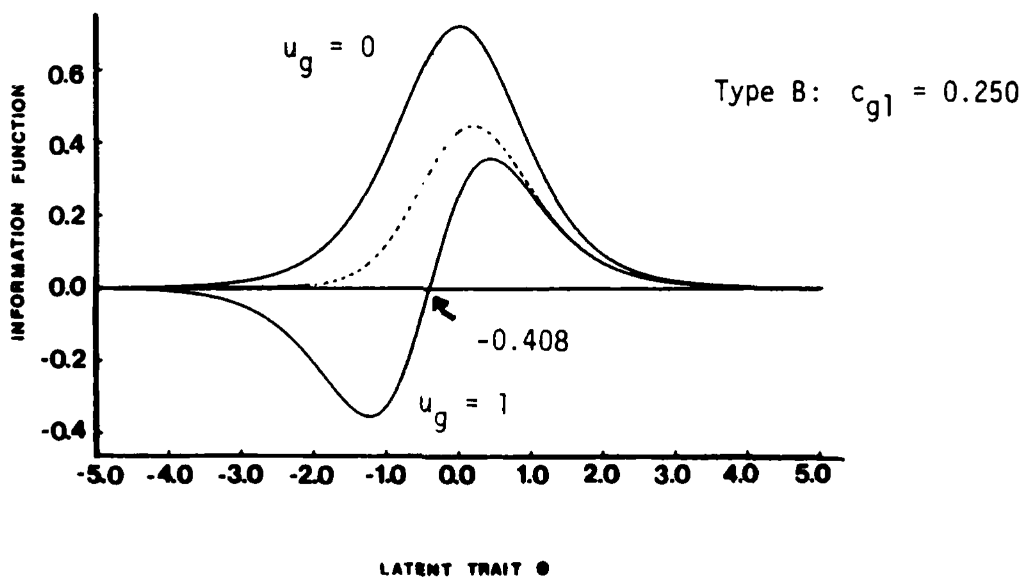


FIGURE 4-2 (Continued)

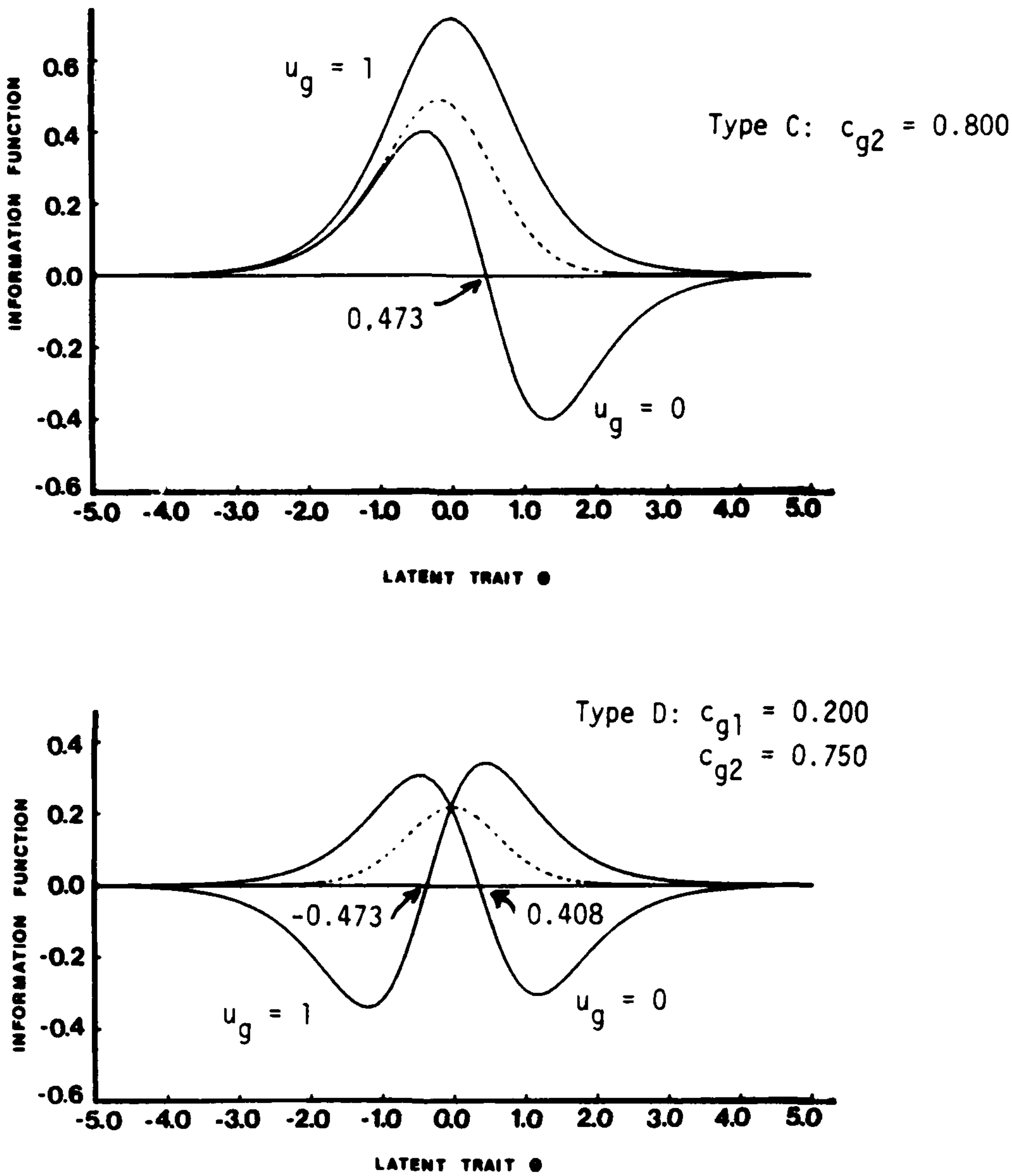


FIGURE 4-2 (Continued)

Since the item information function is the conditional expectation of the item response information function, i.e.,

$$(4.18) \qquad I_g(\theta) = E[I_{u_g}(\theta) \mid \theta] \ ,$$

the values of $\underline{\theta}_g$ and $\overline{\theta}_g$ indicate the direction of the information loss. Figure 4-3 exemplifies the information loss for various values of c_{g1} and/or of c_{g2} for Types B, C and D .

[IV.4] Three-Parameter Logistic Model

Since three-parameter logistic model is one of the most widely used models for dichotomous test items, special observations have been made for this model. This model belongs to Type B, and its item characteristic function is given by

$$(4.19) \qquad P_g(\theta) = c_{g1} + (1 - c_{g1})[1 + exp\{-Da_g(\theta - b_g)\}]^{-1} \ .$$

We also have for the total information Q in this model

$$(4.20) \qquad Q = \pi - 2\tan^{-1}[\frac{c_{g1}}{1 - c_{g1}}]^{1/2} \ .$$

When $c_g = 0.20$, Q equals, approximately, 0.705 π , and when c_g =0.25 , it is approximately 0.667 π .

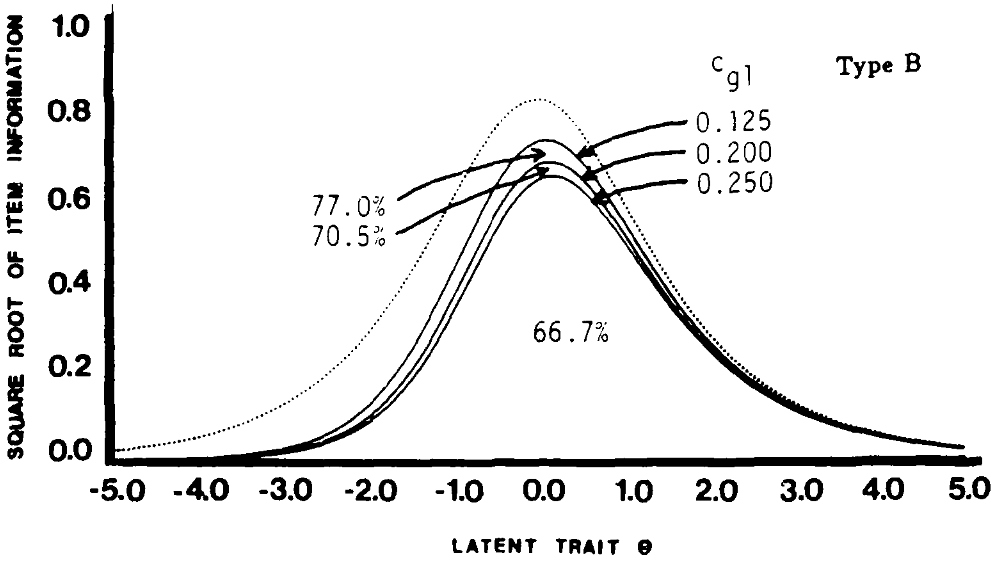
Since four and five are the most commonly used numbers of alternative answers to a multiple-choice test item, the square root of the test information function $I(\theta)$, which is given by

$$(4.21) \qquad [I(\theta)]^{1/2} = [\sum_{g=1}^{n} I_g(\theta)]^{1/2} \ ,$$

was observed for each of the two cases where $c_{g1} = 0.20$ and $c_{g2} = 0.25$, respectively, for eleven sets of different numbers of equivalent items, in comparison with the case where $c_{g1} = 0$, i.e, the (two-parameter) logistic model. These results were also compared with the corresponding results obtained by assuming the three-parameter normal ogive model. Standard error of measurement as a function of ability θ is also observed for these different sets of equivalent items.

[IV.5] Loss in Speed of Convergence of the Conditional Distribution of the Maximum Likelihood Estimate to the Normality

The effect of noise caused by additional parameters, c_{g1} and c_{g2} , is naturally found in the loss of the speed of convergence of the conditional distribution of the maximum likelihood estimate $\hat{\theta}$, given θ , to the normality. This was observed both for different sets of equivalent items as well as those of non-equivalent items. It was discovered that the effect of noise is substantial in decreasing the convergence speed.



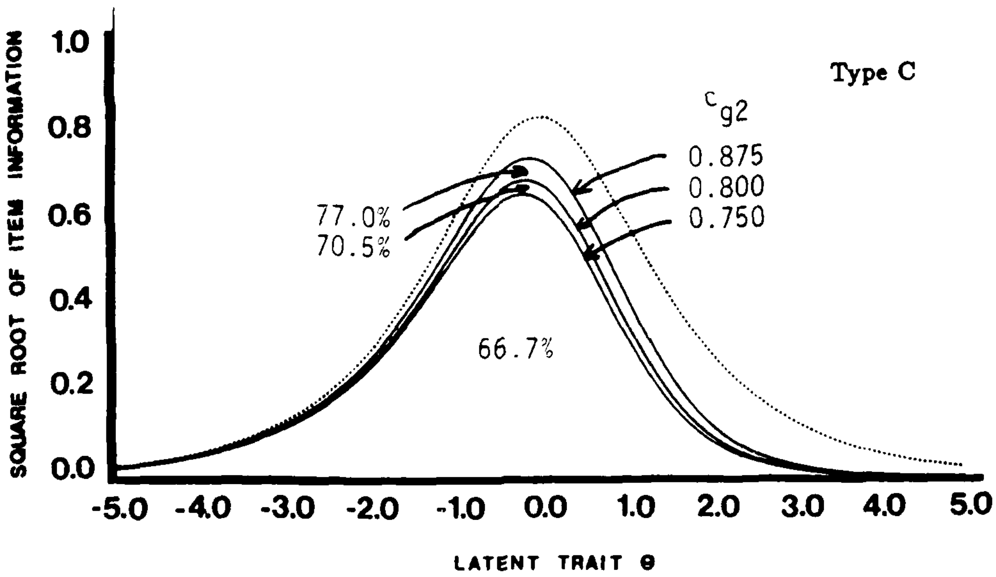


FIGURE 4-3

Square Root of the Item Information Function for Each of Several Hypothetical Items of Types B, C and D (Solid Lines) in Contrast to the One Following the Logistic Model, with $D = 1.7$, $a_g = 1.00$ and $b_g = 0.00$ (Dotted Line). Values of c_{g1} and/or c_{g2} Are as Specified.

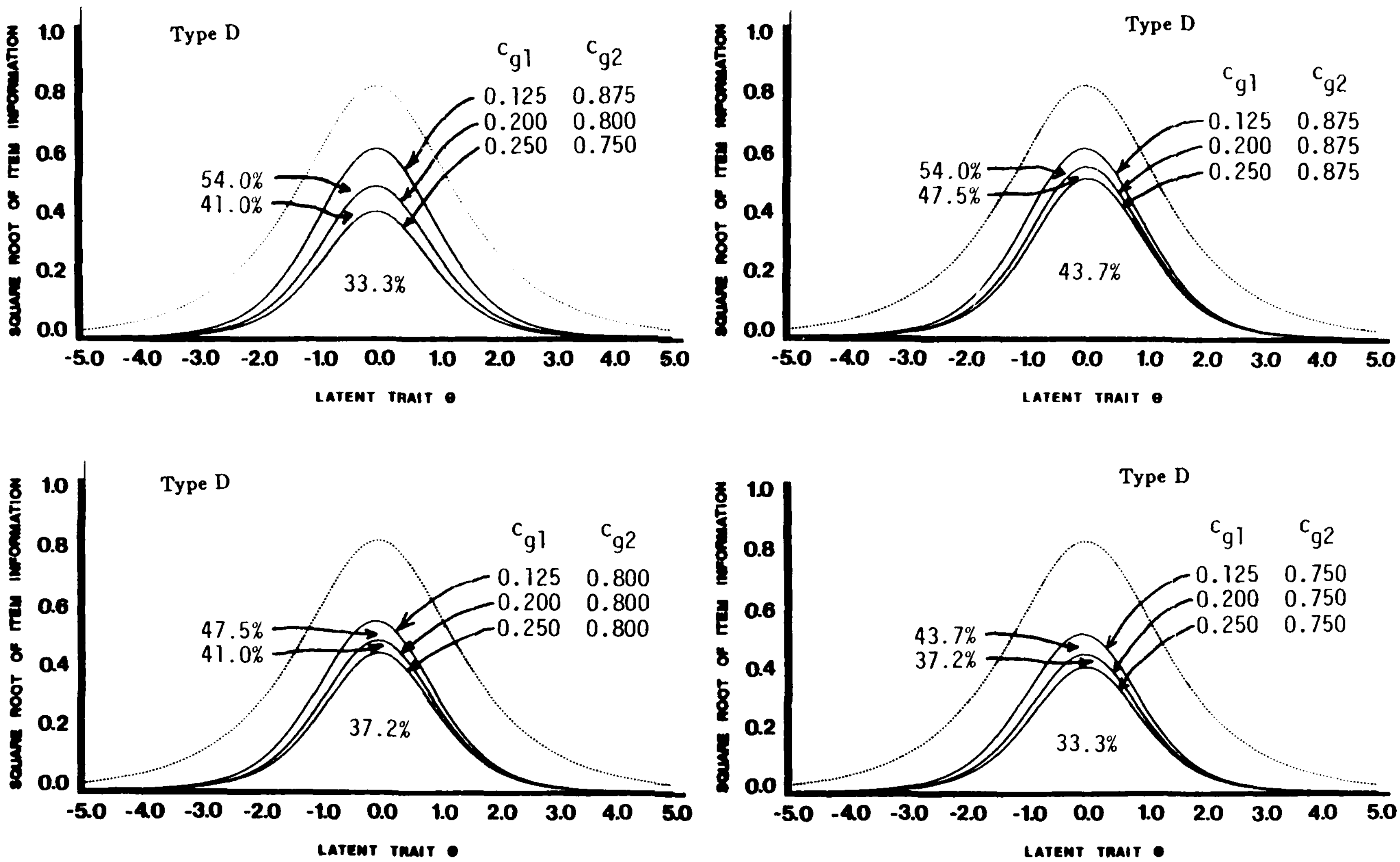


FIGURE 4-3 (Continued)

[IV.6] Discussion

The principal investigator's standpoint is that we should try to eliminate noise by constructing "good" test items, since noise, which may be caused by random guessing, or by some other factors, is nothing but nuisance. Its undesirable effect is probably greater than most researchers think. Because of general indifference and uncritical acceptance of the three-parameter logistic model, however, it seems necessary that someone should quantify the effect of noise incorporated in such models. The effective use of the critical values $\underline{\theta}_g$ and $\overline{\theta}_g$ may be a right step toward the solution.

There are many other developments and observations which are not presented here. They include, among others, observations of the loss of accuracy in ability estimation caused by random guessing in the three-parameter logistic model, both for equivalent items and non-equivalent items.

References

[1] Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick ; *Statistical theories of mental test scores.* Addison-Wesley, 1968, Chapters 17-20.

[2] Lazarsfeld, P. F. Latent structure analysis. In S. Koch (Ed.), *Psychology: a study of a science*, Volume 3. McGraw-Hill, 1959, pages 476-542.

[3] Lord, F. M. and M. R. Novick. *Statistical theories of mental test scores.* Addison-Wesley, 1968.

[4] Samejima, F. Estimation of Latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17, 1969.

[5] Samejima, F. A General model for free-response data. *Psychometrika monograph*, No. 18, 1972.

[6] Samejima, F. Homogeneous case of the continuous response model. *Psychometrika*, 38, 1973a, pages 203-219.

[7] Samejima, F. A comment on Birnbaum's three-parameter logistic model in the latent trait theory. *Psychometrika*, 38, 1973b, pages 221-233.

[8] Samejima, F. Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika*, 39, 1974, pages 111-121.

[9] Samejima, F. A method of estimating item characteristic functions using the maximum likelihood estimate of ability. *Psychometrika*, 42, 1977, pages 163-191.

[10] Samejima, F. Constant information model: a new, promising item characteristic function. *ONR Research Report* 79-1, 1979.

V A Latent Trait Model for Differential Strategies in Cognitive Processes

One of the main objectives of this research project is to "bridge" psychometrics with cognitive psychology through the advancement of latent trait theory. With the rapid progress of microcomputers in the past decade and the accompanied decreases in their cost, many scientific investigations which were considered practically impossible in the past are now within our reach. Thus in many areas of cognitive psychology, where researchers used to conduct their research using relatively small samples of subjects, we can plan our research on a much larger scale. Time is coming, therefore, that latent trait theory will find its way to contribute to the progress of cognitive psychology.

Some cognitive psychologists, who have tried to approach psychometric theories, say that they do not provide them with theories and methods with which they can deal with differential strategies. They are not exactly right, however. As early as in the late nineteen-sixties, the heterogeneous case of the graded reponse level in the context of latent trait theory was proposed (Samejima, 1967) as a model for cognitive processes. Some useful hints for differential strategies are also seen (Samejima, 1972, Section 3.4) under the title "Multi-correct and multi-incorrect responses."

Following the same line, the principal investigator proposed a general latent trait model for differential strategies in cognitive processes, and discussed the topics intrinsic in the model (cf. [I.1.3], [I.1.4]). In this chapter, the outline of these works will be presented.

[V.1] Rationale

The model deals with the unidimensional latent space, in which the latent trait, or "ability", θ assumes any real number. Thus we can write

$$-\infty < \theta < \infty \ . \tag{5.1}$$

Let us take problem solving as an example. Suppose that for solving the problem g we need m_g sequential subprocesses. Let y_g denote the *attainment category* or *attainment score*. One must sucessfully follow all the m_g sequential subprocesses in order to solve the problem g , so the attainment category y_g assumes integers, 0 through $(m_g + 1)$, with $y_g = 0$ indicating that the individual subject has successfully followed none of the subprocesses, and with $y_g = m_g$ meaning that he has completed all m_g subprocesses required to solve the problem. The additional attainment score, $(m_g + 1)$, indicates that the subject has successfully followed the additional subprocess which does not exist but is hypothesized at the end of the entire sequence of subprocesses. Since no one can accomplish this, the conditional probability, given θ , with which the subject obtained the attainment score $(m_g + 1)$ equals zero, regardless of a given value of θ . With this setting, we can see that the general graded response model can readily be applied to the single strategy case of problem solving. Our main objective is, however, to approach a general model for the multiple strategy case, or differential strategies, in the context of latent trait theory.

It is a fairly common phenomenon that there exist more than one way of solving a problem. In proving a mathematical theorem, for example, we often find one proof plus several alternative proofs for one theorem. Figure 5-1 presents a simple example of a two strategy case in the form of a *graph*, each strategy having a small number of subprocesses. In this example, if we take the first strategy to solve the problem, then we must traverse the *path*, $v_0\, v_1\, v_2\, v_3\, v_4$, whereas we must follow another path $u_0\, u_1\, u_2\, u_3\, u_4\, u_5$ if we take the second strategy. (Note that $v_0 = u_0$, $v_1 = u_1$, $v_3 = u_4$ and $v_4 = u_5$.)

When the subject falters, we need additional arcs in the digraph presented as Figure 5-1. Two examples of the directed subgraphs which represent "faltering" are presented in Figure 5-2. These are rather simple examples adding one cycle to each path included in Figure 5-1, making the strategy a *trail* instead of a path. We can conceive of more complex examples, however, in which the subject traverses several cycles repeatedly in a single walk, for example.

In our cognitive process, however, we often choose wrong strategies which do not lead to the solution of the problem at all. Figure 5-3 illustrates such situations in which hollow circles and dashed arcs are added to imply additional paths representing wrong strategies, and two examples of such unsuccessful strategies. Even if the subject took a wrong strategy, he may become aware of his mistake and come back to a previous point in the path and try another strategy. Two examples of such trails are given in Figure 5-4. There are a great many other varieties of paths, trails, and walks, each of which might represent a specified subject's cognitive process. The subject may walk the same cycles over and over again, for example, or he may stop at, say, the vertex v_2 in the path representing the first successful strategy in Figure 5-4 and then may not proceed, and so forth.

It is obvious that following those cycles illustrated in Figure 5-2 and 5-4 will not directly improve the subject's degree of attainment toward the solution of the problem. Thus we can more or less ignore the subject's traversing on cycles, and the things that count are the paths in those graphs, rather than trails or walks which may include one or more cycles. This implies, for instance, that the first trails in Figures 5-2 and 5-4 are treated as equivalent to the completion of Strategy 2, the second trails in those two figures are equivalent to that of Strategy 1, and the two examples of unsuccessful strategies in Figure 5-3 are equivalent to the paths $u_0\, u_1\, u_2$ and $u_0\, u_1\, u_2\, u_3$, respectively.

There is no reason to assume that for a given individual of trait θ the probability of success stays the same when he chooses different successful strategies. Thus the probability with which the subject of trait θ solves the problem using the first successful strategy may not be the same as the one when he uses the second successful strategy, even though the edge $(v_3\, v_4)$ is the same as the edge $(u_4\, u_5)$

[V.2] Differential Strategy Trees

Figure 5-5 presents both the successful and unsuccessful strategies discussed in our example in the form of a tree. Note that not only are the two points v_4 and u_5 in Figure 5-5 the same point in Figure 5-3, but also the two hollow circles marked with * in Figure 5-5 are a single point in Figure 5-3, and so are the two marked with ** . We shall call this kind of tree the *differential strategy tree.* It is a kind of directed graph which contains several paths representing different strategies, joining a common initial endpoint with the distinct other endpoints. We call this initial point a *nothing point,* indicating that, "nothing has been accomplished yet," and the other endpoints for the successful strategies *solution points,* meaning that the "solution has been reached." Since no one can surpass a solution point, it also represents a hypothesized attainment score which no one can obtain.

Figure 5-6 presents a little more complicated example of the digraph and corresponding differential strategy tree, in which only successful strategies are drawn. Thus in our second example, we have five successful strategies and five solution points.

[V.3] A General Model for Differential Strategies

Let w denote the number of successful strategies for solving the probelm g . This number equals the number of solution points in the differential strategy tree, which was illustrated in the preceding section. Each of those w strategies consists of m_{gi} $(i = 1, 2, \ldots, w)$ subprocesses, and they are represented by the vertices, excluding the first and last, both in the digraphs and in the differential strategy trees. In the example which was first presented and discussed in the preceding section, $w = 2$,

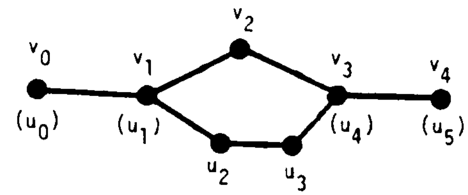


FIGURE 5-1

Example of Two Strategy Case Drawn as a Graph.

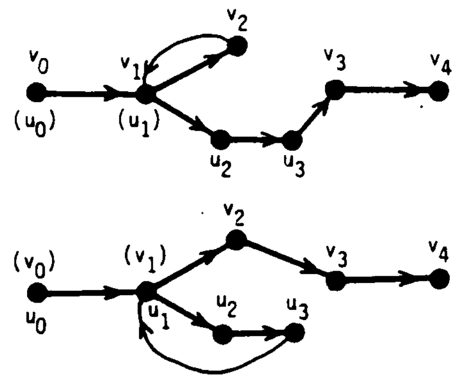


FIGURE 5-2

Two Examples of Directed Subgraphs Representing the Subject's Faltering.

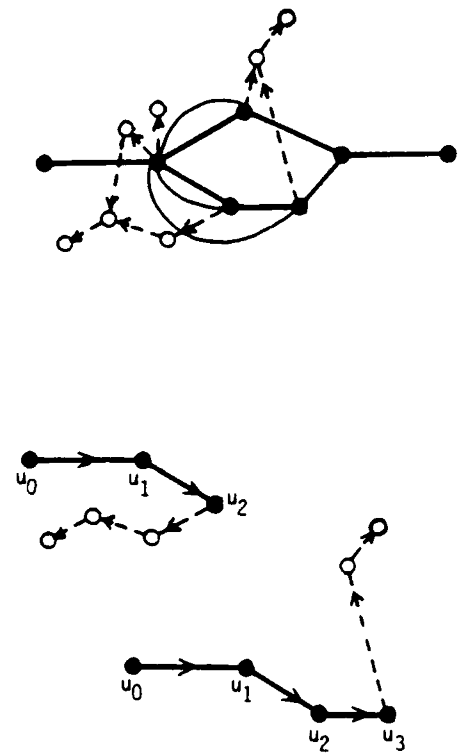


FIGURE 5-3

Diagraph Added by Unsuccessful Strategies (Dashed Lines), and Two Examples of Unsuccessful Strategies

$m_{g1} = 3$ and $m_{g2} = 4$ and the subprocesses are represented by four edges, $(v_0\, v_1), (v_1\, v_2), (v_2\, v_3)$ and $(v_3\, v_4)$, in the path $v_0\, v_1\, v_2\, v_3\, v_4$ representing the first successful strategy, for example. Let y_{gi} be the *attainment score* indicating the degree of attainment of the subject's performance toward the solution of the problem g, which takes on integers 0 through m_{gi} when the subject chooses the strategy i. Figure 5-7 presents the attainment scores assigned to separate edges of the differential strategy tree of our second example.

A general model for differential strategies concerns the assignment of an operating characteristic to each attainment score y_{gi} of each of the w strategies i for solving the problem g. By such an operating characteristic we mean the conditional probability with which the subject of trait θ chooses the strategy i and obtains the attainment score y_{gi}. We notice, however, that in general, if the subject's performance stopped before branching, there is no way to decide which of the two or more strategies he would have taken. For example, $(s_1\, s_2)$ and $(t_1\, t_2)$ in Figure 5-6 are a single edge, and so are $(v_1\, v_2)$ and $(w_1\, w_2)$. Thus we must assign a single operating characteristic for each *edge* of the differential strategy tree. Since each edge represents a union of one or more attainment scores, the operating characteristic is to be assigned to each union. For instance, following an appropriate model, a single operating characteristic will be assigned to the union of $y_{gi} = 0$ for $i = 1, 2, 3, 4, 5$, and the same model will provide us with an operating characteristic solely for $y_{g4} = 3$. For convenience, we shall choose the smallest i in each union, and let y^*_{gsi} denote such a union with s for the actual attainment score. In example 2, for instance, $y^*_{g01} = (y_{g1} = 0) \cup (y_{g2} = 0) \cup (y_{g3} = 0) \cup (y_{g4} = 0) \cup (y_{g5} = 0)$, and $y^*_{g34} = (y_{g4} = 3)$, and none of the unions, y^*_{g02}, y^*_{g03}, y^*_{g04} and y^*_{g05} exists.

Let $M_{y^*_{gs1}}(\theta)$ denote the conditional probability with which the subject of trait θ obtains s as his attainment score in one of the strategies which belongs to y^*_{gsi}, with the joint condition that he has already obtained the score $s - 1$. Since there is no preceding attainment score for $y_{gi} = 0$, and y^*_{g01} is the union of $y_{gi} = 0$ for all the w strategies, the attainment function $M_{y^*_{g01}}(\theta)$ takes on unity throughout the whole range of θ. On the other hand, since $y_{gi} = m_{gi} + 1$ is a hypothesized attainment score which is higher than the full score m_{gi}, the attainment function $M_{y^*_{g(m_{gi}+1)i}}(\theta)$ assumes zero for the entire range of θ for each of the w strategies. Thus we can write

$$(5.2) \qquad M_{y^*_{gsi}}(\theta) \begin{cases} = 1 & i = 1, \qquad s = 0 \\ = 0 & i = 1, 2, \ldots, w, \qquad s = m_{gi} + 1 . \end{cases}$$

Note that in (5.2) the first line indicates a single function for the union of $y_{gi} = 0$ for $i = 1, 2, ..., w$, while the second line indicates w separate functions for $i = 1, 2, ..., w$.

Hereafter, we shall assume that each attainment function $M_{y^*_{gsi}}(\theta)$ is three-times differentiable with respect to θ. Note that this assumption does not contradict (5.2).

Let $P^*_{y^*_{gsi}}(\theta)$ be the conditional probability assigned to the union of attainment scores y^*_{gsi}, with which the subject of trait θ chooses a strategy which belongs to y^*_{gsi} and obtains the attainment score s or greater. We shall call this function the ***cumulative operating characteristic of the attainment score union*** y^*_{gsi}.

From the definitions of this function and the attainment function $M^*_{gsi}(\theta)$, we can write

$$(5.3) \qquad P^*_{y^*_{gsi}}(\theta) = \prod_{k=0,*i}^{s} M_{y^*_{gk(*i)}}(\theta)$$

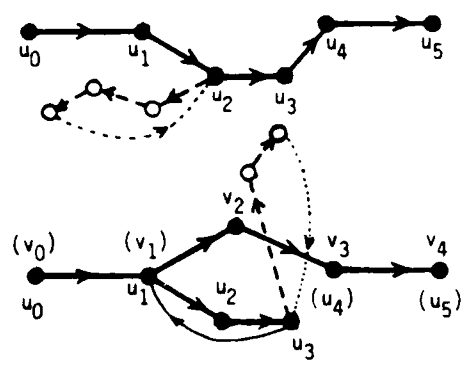


FIGURE 5-4

Two Trails Representing the Subject's Faltering When He Has Taken Unsuccessful Strategies.

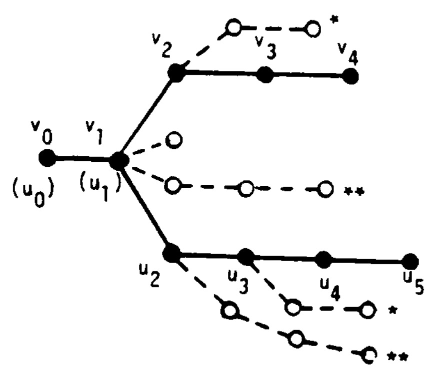


FIGURE 5-5

Differential Strategy Tree Including Both Successful and Unsuccessful Strategies.

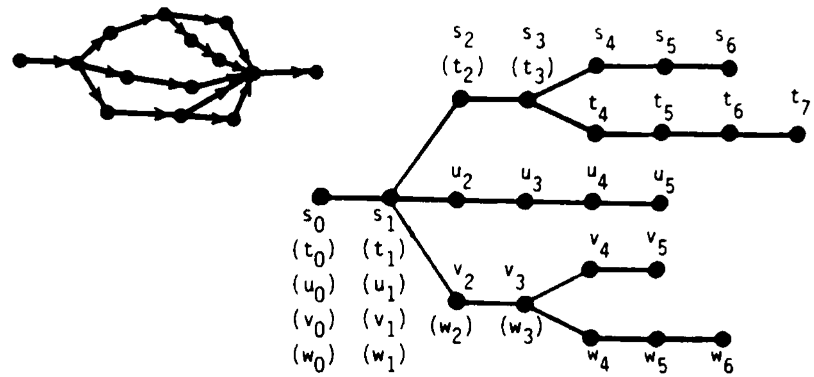


FIGURE 5-6

Another Example of Diagraph Representing Five Successful Strategies and Its Differential Strategies Tree.

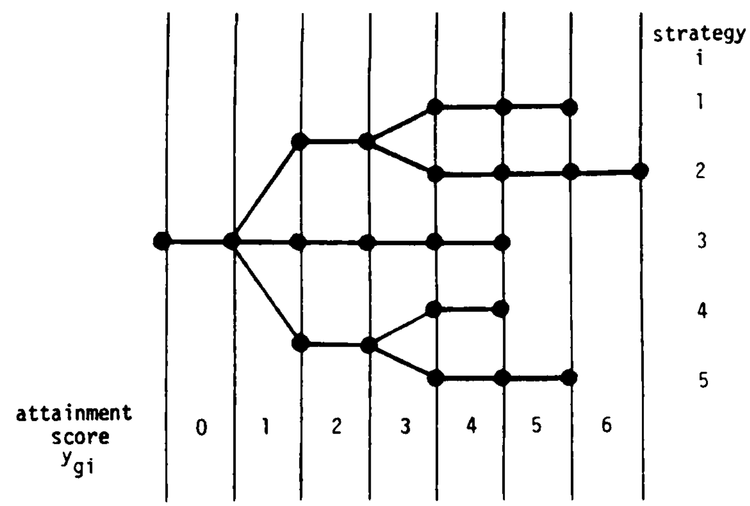


FIGURE 5-7

Examples of Attainment Scores Given to Different Subgraphs in Example 1, and Those Assigned to Separate Lines of the Differential Strategy Tree in Example 2.

where $*i$ indicates the closest integer less than or equal to i for which the union of attainment scores exists. In particiular, we have

$$(5.4) \qquad P^*_{y^*_{gsi}}(\theta) \begin{cases} = 1 & s = 0 \\ = M_{y^*_{gsi}}(\theta) & s = 1 \\ = 0 & s = m_{gi} + 1 \ . \end{cases}$$

Note that the first line of (5.4) indicates a single function for the union of $y_{gi} = 0$ for $i = 1,2,\ldots,w$, the second line indicates one or more functions depending upon the branching, and the third line indicates the w separate functions for $i = 1,2,\ldots,w$. In Example 1, the second line includes two functions, while in Example 2 it includes three functions.

Let $A^*_{y^*_{gsi}}(\theta)$ be the first derivative or the natural logarithm of the cumulative operating characteristic $P^*_{y^*_{gsi}}(\theta)$, that is,

$$(5.5) \qquad A^*_{y^*_{gsi}}(\theta) = \frac{\partial}{\partial\theta} \log P^*_{y^*_{gsi}}(\theta) \ .$$

Note that this function is unchanged when $P^*_{y^*_{gsi}}(\theta)$ is multiplied by a constant. To be more specific, let $\Psi(\theta)$ be a function defined by

$$(5.6) \qquad P^*_{y^*_{gsi}}(\theta) = c_1 + (c_2 - c_1)\Psi(\theta) \ ,$$

where $0 \le c_1 \le c_2 \le 1$. The firt derivative of the natural logarithm of $\frac{\partial}{\partial\theta} \log P^*_{y^*_{gsi}}(\theta)$ is given by $\frac{\partial}{\partial\theta} \log[c_1 + (c_2 - c_1)\Psi(\theta)]$, which equals $\frac{\partial}{\partial\theta} \log \Psi(\theta)$ if, and only if, $c_1 = 0$. The formula (5.6) has been observed in a somewhat different context (cf. [I.1.1]) and these observation was summarized in Chapter IV, where $P^*_{y^*_{gsi}}(\theta)$ is replaced by any strictly increasing function of θ with zero and unity as its two asymptotes. We have called the four different types of functions derived from (5.6) Types A, B, C and D (cf. Chapter IV), depending upon the values of c_1 and c_2 , i.e., the function is of Type A when $0 = c_1 < c_2 = 1$; of Type B when $0 < c_1 < c_2 = 1$; of Type C when $0 = c_1 < c_2 < 1$; and of Type D when $0 < c_1 < c_2 < 1$, respectively. This implies that the cumulative operating characteristics of Types A and C may share the same function for $A^*_{y^*_{gsi}}(\theta)$, and we can say the same for those of Types B and D. The necessary and sufficient condition that $M_{y^*_{gsi}}(\theta)$ be strictly increasing in θ is that the inequality

$$(5.7) \qquad A_{y^*_{gsi}}(\theta) > A_{y^*_{g(s-1)(*i)}}(\theta) \qquad s = 1,2,\ldots,(m_{gi}+1)$$

holds almost everywhere with respect to θ .

The operating characteristic, $P^*_{y^*_{gsi}}(\theta)$, defined for the union of the attainment scores y^*_{gsi} is given by

$$P_{y^*_{gsi}}(\theta) = P^*_{y^*_{gsi}}(\theta) - \sum_{j*} P^*_{y^*_{g(s+1)j}}(\theta) \ , \tag{5.8}$$

where $\sum_{j*}$ indicates the summation over all the strategies j branching from the point which lies immediately after the line representing y^*_{gsi} .

This operating characteristic can be considered as the likelihood function in estimating the subject's latent trait θ .

When there are more than one problem to solve, i.e., $g = 1, 2, \ldots, n$, satisfying the conditional independence of the attainment scores across the different items, given θ , the maximum likelihood estimation of the subject's latent trait can be performed on the basis of the response pattern V , such that

$$V' = (y_{1i_1}, y_{2i_2}, \cdots, y_{gi_g}, \cdots, y_{ni_n}) \tag{5.9}$$

for the n problem solving tasks, where i_g is a strategy for solving the problem g and y_{gi_g} is the attainment score when the subject chooses the strategy i_g for solving the problem g . Let $P_V(\theta)$ be the operating characteristic of the specific response pattern V . We can write

$$P_V(\theta) = \prod_{*} P_{y^*_{gsi}}(\theta) \ , \tag{5.10}$$

where $\prod_*$ indicates the multiplication over every union y^*_{gsi} to which an element of V belongs.

It is beneficial to search for a family of models which provide us with a unique maximum for every possible response pattern given by (5.9). This can be done as a generalization of the unique maximum condition proposed for the graded response model (cf. Samejima, 1969, 1972).

The *basic function*, $A_{y^*_{gsi}}(\theta)$, for the union of attainment scores y^*_{gsi} is defined by

$$A_{y^*_{gsi}}(\theta) = \frac{\partial}{\partial\theta} \log P_{y^*_{gsi}}(\theta) \ . \tag{5.11}$$

The maximum likelihood estimate, $\hat{\theta}_V$, of the subject's latent trait based upon his response pattern is given as the solution of the likelihood equation such that

$$\begin{aligned} \frac{\partial}{\partial\theta} \log P_V(\theta) &= \sum_{*} \frac{\partial}{\partial\theta} \log \ P_{y^*_{gsi}}(\theta) \\ &= \sum_{*} A_{y^*_{gsi}}(\theta) \equiv 0 \ , \end{aligned} \tag{5.12}$$

where $\sum_*$ indicates the summation over every union y^*_{gsi} to which an element of V belongs. A sufficient condition that a unique modal point exists for the likelihood function $P_V(\theta)$ of each and every response pattern V is that this basic function is strictly decreasing in θ with non-negative and non-positive values as its two asymptotes, respectively, for every union y^*_{gsi} . This can be shown in the same way that we did for the basic function $A_{x_g}(\theta)$ of the graded item score x_g (cf. Samejima, 1969). For brevity, sometimes we call this condition the unique maximum condition.

Similarities between the differential strategies in problem solving and the multi-correct responses in testing are obvious. If we consider two or more different strategies which lead to the solution of the problem as two or more different answers to a question, then they will be treated as multi-correct responses. We can see that the concept of multi-correct responses can be transfered to differential strategies, when there exist more than one successful strategy in solving a problem.

[V.4] Homogeneous Case

The Homogeneous case of the graded response level has been developed and discussed (Samejima, 1972) as a generalization of a family of models on the dichotomous response level. Sufficient conditions that a model provides us with a unique modal point for the likelihood function of each and every response pattern have been investigated. In the homogeneous case, a sufficient condition is that, for an arbitrary item score $x_g \ (\neq 0)$, the cumulative operating characteristic $P^*_{x_g}(\theta)$ is of Type A, i.e., strictly increasing in θ with zero and unity as its two asymptotes, and its asymptotic basic function, $\tilde{A}_{xg}(\theta)$, which is defined by

$$\tilde{A}_{x_g}(\theta) \ = \ \frac{\partial}{\partial\theta}[\log\{\frac{\partial}{\partial\theta}P^*_{x_g}(\theta)\}] \ , \tag{5.13}$$

is strictly decreasing in θ . The satisfaction of this sufficient condition also implies two desirable features of the model such that: 1) the operating characteristic of each graded item score of each item has a single modal point, and 2) those modal points for a single item are arranged in the same order as the item scores themselves. The normal ogive and logistic models, which have been generalized from the corresponding models on the dichotomous response level, are two examples of the models which satisfy the above sufficient condition.

These models of the homogeneous case on the graded response level can be generalized to provide us with those which belong to the general model of differential strategies. Let $\Psi(\theta)$ be a function of Type A. We shall consider the cumulative operating characteristic, $P^*_{y^*_{gsi}}(\theta)$, of the union of attainment categories y^*_{gsi} such that

$$P^*_{y^*_{gsi}}(\theta) \ = \ \beta_{y^*_{gsi}} \ \Psi(\theta - \alpha_{y^*_{gsi}}) \ , \tag{5.14}$$

where $\alpha_{y^*_{g0i}}$ is negative infinity, $\alpha_{y^*_{g(m_{gi}+1)i}}$ is positive infinity for $i = 1, 2, \ldots, w$, and the values of $\alpha_{y^*_{gsi}}$ are ordered in the same way as those of s for every strategy, and $\beta_{y^*_{gsi}}$ is a constant which equals unity for $s \ = \ 0$ and in general satisfies

$$\sum_{j_*} \beta_{y^*_{gsj}} \ = \ \beta_{y^*_{g(s-1)i}} \ , \tag{5.15}$$

with $\sum_{j^*}$ indicating the summation over all the strategies j branching from the point of the differential strategy tree which is located right after the line representing the union $y^*_{g(s-1)i}$. From (5.15) it is obvious that, as far as there is no branching, $\beta_{y^*_{gsi}} \ = \ \beta_{y_{g(s-1)i}}$.

A sufficient condition that the model satisfies the unique maximum condition is: 1) that the values of the constant $\alpha_{y^*_{gsj}}$ are the same for all the strategies j which branch from the vertex located immediately after the edge representing y^*_{gsi} , and 2) that we have

$$(5.16) \qquad \frac{\partial}{\partial\theta}[\log\{\frac{\partial}{\partial\theta}\Psi(\theta)\}] \ < \ 0$$

almost everywhere in the domain of θ . To prove this, we obtain from (5.8), (5.14), (5.15) and the definition of the basic function $A_{y^*_{gsi}}(\theta)$, which was given by (5.11),

$$(5.17) \qquad \begin{aligned} A_{y^*_{gsi}}(\theta) &= \frac{\partial}{\partial\theta}\log P_{gsi}(\theta) \\ &= \frac{\partial}{\partial\theta}\log[\beta_{y^*_{gsi}}\Psi(\theta-\alpha_{y^*_{gsi}}) - \sum_{j*}\beta_{y^*_{g(s+1)j}}\Psi(\theta-\alpha_{y^*_{g(s+1)j}})] \ . \end{aligned}$$

where $\sum_{j_*}$ indicates the summation over all the strategies j branching from the vertex which lies immediately after the line representing the union y^*_{gsi} . By virtue of the first condition, we can rewrite (5.17) in the form

$$(5.18) \qquad \begin{aligned} A_{y^*_{gsi}}(\theta) &= \frac{\partial}{\partial\theta}\log[\beta_{y^*_{gsi}}\{\Psi(\theta-\alpha_{y^*_{gsi}}) - \Psi(\theta-\alpha_{y^*_{g(s+1)j}})\}] \\ &= \frac{\partial}{\partial\theta}\log\{\Psi(\theta-\alpha_{y^*_{gsi}}) - \Psi(\theta-\alpha_{y^*_{g(s+1)j}})\} \ . \end{aligned}$$

We notice that, if we replace y^*_{gsi} by the graded item score x_g and use $\Psi(\theta-\alpha_{x_g})$ as the cumulative operating characteristic $P^*_{x_g}(\theta)$, the last form of (5.18) is identical with the basic function of the graded item score, and the left hand side of (5.16) is identical with the corresponding asymptotic basic function. Thus we can say that all the unions, y^*_{gsi} , are equivalent to syndrome response categories (cf. Samejima, 1972, Section 5.2), and a unique maximum is assured for every possible response pattern.

If, for example, $\Psi(\theta)$ is a normal ogive function or a logistic distribution function, then (5.16) is satisifed (Samejima, 1972, Section 5.2), and we can develop the normal ogive model and the logistic model in the context of the general model for differential strategies, and both of them satisfy the unique maximum condition. In these two models, the cumulative operating characteristics are defined by

$$(5.19) \qquad P^*_{y^*_{gsi}}(\theta) \ = \ \beta_{y^*_{gsi}}(2\pi)^{-1/2}\int_{-\infty}^{a_g(\theta-b_{y^*_{gsi}})} e^{-u^2/2}\,du$$

and

$$(5.20) \qquad P^*_{y^*_{gsi}}(\theta) \ = \ \beta_{y^*_{gsi}}[1+exp\{-Da_g(\theta-b_{y^*_{gsi}})\}]^{-1} \ ,$$

respectively, where a_g (>0) is the discrimination parameter specific for each problem g , $b_{y^*_{gsi}}$ is a difficulty parameter defined for each union of attainment scores, with $b_{y^*_{g0i}} = -\infty$ and $b_{y^*_{g(m_{gi}+1)i}} = \infty$ and all those values are arranged in the same order as s with respect to each strategy, and D in (5.20) is a scaling factor which assumes 1.7 to retain the same set of parameter values as those in the normal ogive model.

Figure 5-8 presents the set of ten cumulative operating characteristics $P^*_{y^*_{gsi}}(\theta)$ in Example 1, with the parameter values such that $a_g = 1.00$, $b_{y^*_{g11}} = b_{y^*_{g12}} = -2.50$, $b_{y^*_{g21}} = -1.00$, $b_{y^*_{g31}} = 0.50$,

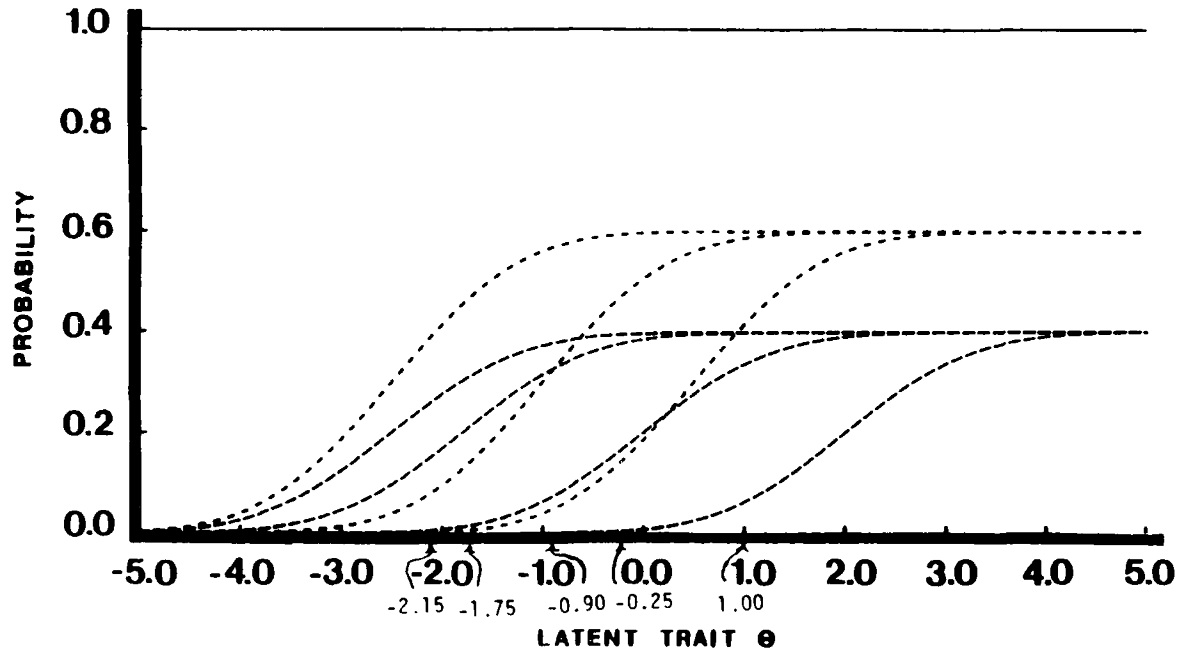


FIGURE 5-8

Cumulative Operating Characteristic of the Union of $(y_{g1} = 0)$ and $(y_{g2} = 0)$ (Solid Line), Those of $(y_{g1} = 1)$, $(y_{g1} = 2)$ and $(y_{g1} = 3)$, Respectively (Dotted Lines), and Those of $(y_{g2} = 1)$, $(y_{g2} = 2)$, $(y_{g2} = 3)$ and $(y_{g2} = 4)$, Respectively (Dashed Lines).

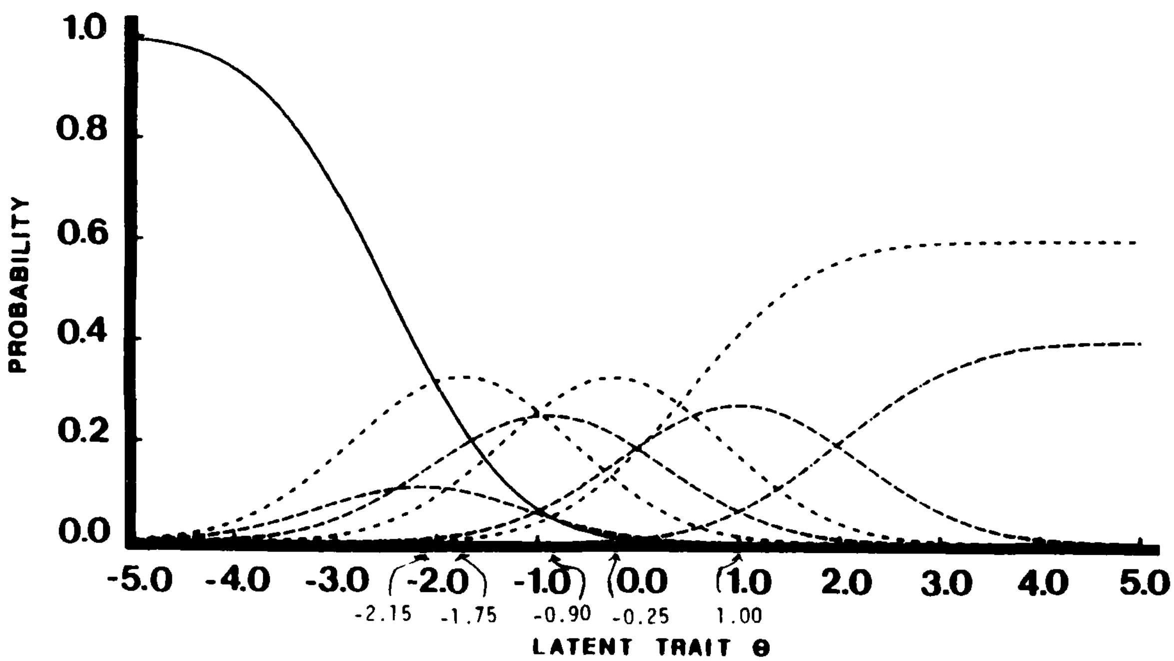


FIGURE 5-9

Operating Characteristic of the Union of $(y_{g1} = 0)$ and $(y_{g2} = 0)$ (Solid Line), Those of $(y_{g1} = 1)$, $(y_{g1} = 2)$ and $(y_{g1} = 3)$, Respectively (Dotted Lines), and Those of $(y_{g2} = 1)$, $(y_{g2} = 2)$, $(y_{g2} = 3)$ and $(y_{g2} = 4)$, Respectively (Dashed Lines).

$b_{y^*_{g22}}=-1.80$, $b_{y^*_{g32}} = 0.00$, $b_{y^*_{g43}} = 2.00$, $\beta_{y^*_{g11}} = 0.60$ and $\beta_{y^*_{g12}} = 0.40$. The corresponding operating characteristics are drawn in Figure 5-9, in which all the modal points except for the negative and positive infinities are shown.

[V.5] Single Strategy Case

In the single strategy case where there is only one successful strategy in our problem solving, things are much more simplified. There exists a parallelism with the graded response model (Samejima, 1972), with the replacement of the item score x_g by the attainment score y_g of the unique successful strategy.

Let $A^*_{y_g}(\theta)$ be the first partial derivative of the natural logarithm of $P^*_{y_g}(\theta)$, where $P^*_{y_g}(\theta)$ equals $P^*_{y^*_{gsi}}(\theta)$ with the replacement of y^*_{gsi} by the single set of attainment score y_g , such that

(5.21) $$A^*_{y_g}(\theta) = \frac{\partial}{\partial\theta} \log P^*_{y_g}(\theta) \qquad y_g = 0, 1, \ldots, (m_g + 1) \ .$$

It has been shown that the necessary and sufficient condition that $M_{y_g}(\theta)$ be strictly increasing in θ is that the inequality

(5.22) $$A^*_{y_g}(\theta) > A^*_{(y_g-1)}(\theta) \qquad y_g = 1, 2, \ldots, (m_g + 1)$$

holds almost everywhere with respect to θ (cf. Samejima, 1967, 1972).

In the homogeneous case $P^*_{y_g}(\theta)$ has zero and unity as its two asymptotes for $y_g = 1, 2, \ldots, m_g$ and, furthermore, we can write

(5.23) $$P^*_s(\theta) = P^*_r(\theta - \alpha_{rs}) \ ,$$

where r and s are two arbitrarily selected attainment categories with $r < s$, and α_{rs} is a positive finite constant. We obtain from (5.21) and (5.23)

(5.24) $$A^*_s(\theta) = A^*_r(\theta - \alpha_{rs}) \ .$$

From (5.22) and (5.24) it is obvious that a sufficient, though not necessary, condition that $M_{y_g}(\theta)$ be strictly increasing in θ for $y_g = 1, 2, \ldots, m_g$ is that $A^*_{y_g}(\theta)$ is strictly decreasing in θ for an arbitrarily chosen attainment category out of 1 through m_g . When m_g tends to positive infinity, and α_{rs} for two adjacent attainment categories tends to zero, in the limiting situation this condition becomes the necessary and sufficient condition, for it requires that

(5.25) $$A^*_{y_g}(\theta) > A^*_{y_g}(\theta + \varepsilon)$$

for any small positive value of ε . Note that this condition is satisfied whenever the unique maximum condition is satisfied. Above all, when the asymptotic basic function, $\tilde{A}_{y_g^*}(\theta)$, which is defined by

$$(5.26) \qquad \tilde{A}^*_{y_g}(\theta) = [\frac{\partial^2}{\partial\theta^2} P^*_{y_g}(\theta)][\frac{\partial}{\partial\theta} P^*_{y_g}(\theta)]^{-1} \ ,$$

is strictly decreasing in θ , not only $M_{y_g}(\theta)$ of each subprocess is strictly increasing in θ , however finely differentiated it may be, but also a unique maximum is assured for the likelihood function of each and every possible response pattern which consists of such attainment scores of different tasks (cf. Samejima, 1972, Sections 5.1 and 5.2). It has been shown (Samejima, 1967, 1972), for example, that in the normal ogive model and in the logistic model on the graded response level (Samejima, 1969) this condition is satisfied. In the former example, we can write

$$(5.27) \qquad P^*_{y_g}(\theta) = (2\pi)^{-1/2} \int_{-\infty}^{a_g(\theta - b_{y_g})} exp(-u^2/2)\, du$$

and in the latter

$$(5.28) \qquad P^*_{y_g}(\theta) = [1 + exp\{-Da_g(\theta - b_{y_g})\}]^{-1} \ ,$$

where a_g is the item discrimination parameter and b_{y_g} is the item difficulty parameter for the attainment category y_g , and D is a scaling factor which is usually set equal to 1.7 (Birnbaum, 1968). In both models, the upper asymptote of $M_{y_g}(\theta)$ for $y_g = 1, 2, \ldots, m_g$ is unity, while the lower asymptote is zero in the normal ogive model and $exp[-Da_g(b_{y_g} - b_{(y_g-1)})]$ in the logistic model. This lower asymptote in the logistic model depends upon the distance between the difficulty parameters of the two adjacent attainment categories, assuming zero for $y_g = 1$ and positive numbers less than unity otherwise. In both models, $M_{y_g}(\theta)$ for $y_g = 2, 3, \ldots, m_g$ tends to unity for the entire range of θ as b_{y_g} approaches $b_{(y_g-1)}$, and tends to $P^*_{y_g}(\theta)$ as b_{y_g} departs from $b_{(y_g-1)}$ (cf. Samejima, 1972, Figure 5-2-1).

This simplified model will be useful not only for cognitive processes, but also for paper-and-pencil testing, provided that the test is constructed in such a way that each item includes only one successful path to reach the correct answer. Such an example is given in the research report [I.1.3].

[V.6] Information Provided by Differential Strategies

The *information function*, $I_{y^*_{gsi}}(\theta)$, for the union of the attainment scores y^*_{gsi} is defined by

$$(5.29) \qquad I_{y^*_{gsi}}(\theta) = -\frac{\partial^2}{\partial\theta^2} \log P_{y^*_{gsi}}(\theta) = -\frac{\partial}{\partial\theta} A_{y^*_{gsi}}(\theta) \ ,$$

where $P_{y^*_{gsi}}(\theta)$ is the operating characteristic, and $A_{y^*_{gsi}}(\theta)$ is the basic function, of y^*_{gsi} , respectively. This function is non-negative whenever the unique maximum condition is satisfied. In the homogeneous case, if there is a single value $\alpha_{y^*_{gsj}}$ common for all the strategies j , which leads to the satisfaction of the unique maximum condition, then we can write

$$(5.30) \qquad I_{y^*_{gsi}}(\theta) = -\frac{\partial^2}{\partial\theta^2} \log\{\Psi(\theta - \alpha_{y^*_{gsi}}) - \Psi(\theta - \alpha_{y^*_{g(s+1)i}})\} \ .$$

For the item information function $I_g(\theta)$ we can write

$$(5.31) \qquad I_g(\theta) = E[I_{y^*_{gs}}(\theta) \mid \theta] = \sum_{y^*_{gs}} I_{y^*_{gs}}(\theta) P_{y^*_{gs}}(\theta)$$

where $\sum_{y^*_{gs}}$ indicates the summation over all the unions of attainment scores, or over all the edges in the differential strategy tree. It is obvious that, in general, the more subprocesses we have within each strategy the greater amount of item information we get, with the continuous subprocesses as the limiting case. The differentiation of strategies itself does not necessarily increase the amount of item information, however.

When we have n problem solving tasks which require the same latent trait θ , for the test information function $I(\theta)$ we can write

$$(5.32) \qquad I(\theta) = \sum_{g=1}^{n} I_g(\theta) \; .$$

[V.7] Discussion

A question may arise as to which estimate of the latent trait should be taken if the subject faltered from one strategy to another and did not reach the solution of the problem. One answer to this question may be to take the attainment score of the strategy that he took last, and use its corresponding operating characteristic in estimating his latent trait. Another answer may be to compare the resultant estimates of θ obtained by the separate strategies the subject has taken and select the highest estimate.

The usefulness of the proposed model is yet to discover. We need the collabortion of cognitive psychologists who are willing to collect data on larger samples, taking advantage of modern technologies.

References

[1] Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick; *Statistical theories of mental test scores.* Addison-Wesley, 1968, Chapters 17-20.

[2] Indow, T., and F. Samejima. *LIS Measurement scale for non- verbal reasoning ability.* (With four tests, LIS-P, LIS-C, LIS-S and LIS-U). Tokyo: Nippon Bunka Kagakusha, 1962.

[3] Indow, T. and F. Samejima. On results obtained by the absolute scaling model and the Lord model in the field of intelligence. *Monographs of the Psychological Laboratory on the Hiyoshi Campus*, Keio University, 1966, No. 3 (in English).

[4] Samejima, F. A general model for the operating characteristic of graded item response. Chapel Hill, North Carolina: *The L. L. Thurstone Psychometric Laboratory Report*, 55, 1967.

[5] Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17, 1969.

[6] Samejima, F. A general model for free-response date. *Psychometrika monograph*, No. 18, 1972.

VI Latent Trait Models for Partially Continuous and Partially Discrete Responses

A set of latent trait models were proposed during this research period, which deals with the mixture of continuous and discrete responses. This family of models is an expansion and a generalization of the one proposed by the principal investigator in 1973 (Samejima, 1973), in which the open response situation is dealt with. The family is represented by the closed response situation, and it also includes the model for the open response situation as a special case, as well as those models for the open/closed and the closed/open response situations.

In this chapter, the outline of these new models will be described, and one separate ongoing research project on the Rorschach diagnosis for which these models are to be used will be introduced as an example. For the details and further information about the models, see [I.1.5].

[VI.1] Rationale

Let θ be the *unidimensional latent trait*, or any hypothetical construct, which assumes all real numbers. Let g $(= 1, 2, \ldots, n)$ be an *item*, which is the smallest, concrete entity devised for measuring the latent trait. The assumption that our latent space is unidimensional implies that the conditional or local independence of the distributions of the item responses of separate items, given θ, holds in the unidimensional latent space.

Distinction between the open response situation and the closed response situation may be well illustrated, schematically, by Figure 6-1. Suppose that the subject is asked to check a point on a line segment illustrated in Figure 6-1 in accordance with his judgment required for the task in item g. Without loss of generality, we can assign the item score z_g which varies zero through unity for each point on the line segment.

It will be reasonable to assume that the probability assigned to any particular point on the line segment is nil, provided that the subject is not allowed to check either of the two endpoints. We call it the open response situation, and assumes a continuous distribution for the item score z_g. If the subject is allowed to check either of the two endpoints as well as the others, however, the probability assigned to these points may not be nil. We call this second situation the closed response situation, and the distribution of z_g must be discrete at the two endpoints, i.e., at $z_g = 0$ and $z_g = 1$, and continuous otherwise. In similar manners we can define the open/closed response situation and the closed/open response situation.

[VI.2] Conditional Distribution of the Item Score

Let $P^*_{z_g}(\theta)$ be the conditional probability with which the subject obtains the item score z_g or greater, given θ. A general mathematical form for $P^*_{z_g}(\theta)$ in the homogeneous case of the continuous response model (Samejima, 1973) is given by

$$(6.1) \qquad P^*_{z_g}(\theta) = \int_{-\infty}^{a_g(\theta - b_{z_g})} \Psi_g(t)dt \ ,$$

with

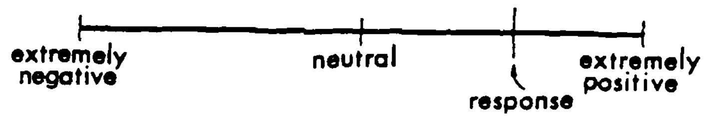


FIGURE 6-1

An Example of the Response Formats Which Allow Continuous Responses.

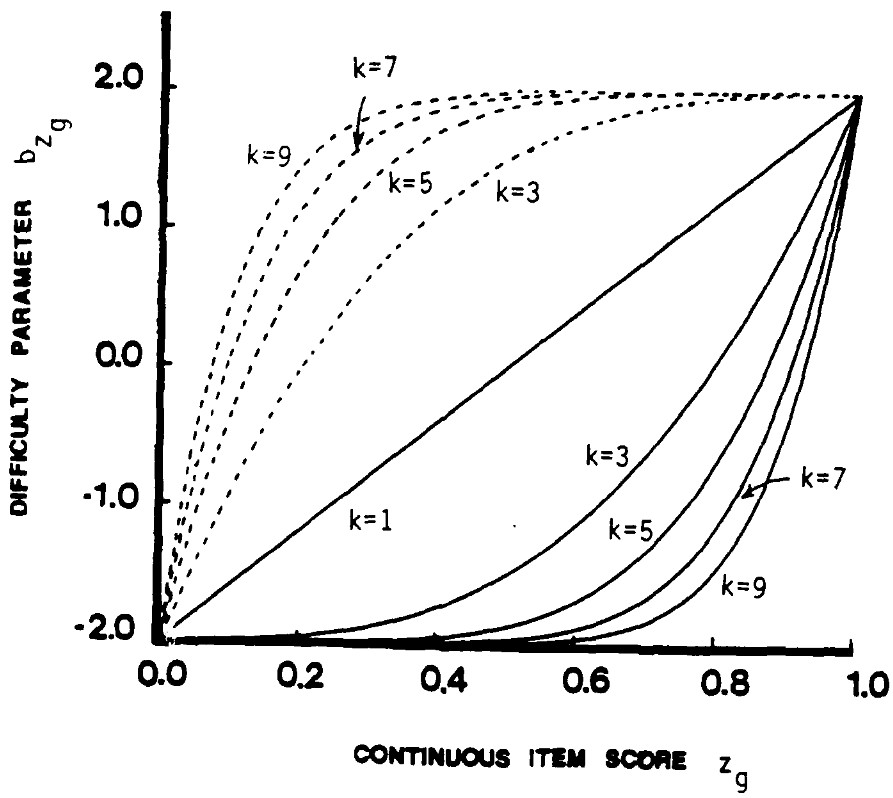


FIGURE 6-2

Five Hypothetical Functional Relationships (Solid Lines) between the Continuous Item Score z_g and the Difficulty Parameter b_{z_g} , Which Are Given by $b_{z_g} = b_0 + (b_1 - b_0)z_g^k$, for k = 1,3,5,7,9 , with the Parameters, $b_0 = -2.0$ and $b_1 = 2.0$, and the Corresponding Relationships (Dotted Lines) Given by $b_{z_g} = b_1 - (b_1 - b_0)(1 - z_g)^k$. Closed Response Situation.

(6.2)
$$\begin{cases} \lim_{\theta \to -\infty} P^*_{z_g}(\theta) = 0 \\ \lim_{\theta \to \infty} P^*_{z_g}(\theta) = 1 \end{cases} ,$$

where $a_g(> 0)$ is the item discrimination parameter, b_{z_g} is the item response difficulty parameter, and $\Psi_g(.)$ is a specific continuous function which characterizes the model, and is positive almost everywhere. Two examples of $\Psi_g(t)$ are the normal ogive function and the logistic function, which are specified, respectively, by

(6.3)
$$\Psi_g(t) = (2\pi)^{-1/2} exp(-t^2/2) ,$$

and

(6.4)
$$\Psi_g(t) = D\, exp(-Dt)\, [1 + exp(-Dt)]^{-2} ,$$

where D is a scaling factor. The operating density characteristic, $H_{z_g}(\theta)$, has been defined, and it can be written in the form

(6.5)
$$H_{z_g}(\theta) = a_g \Psi_g\{a_g(\theta - b_{z_g})\}[\tfrac{d}{dz_g} b_{z_g}] \qquad 0 < z_g < 1 .$$

Let $P_{z_g}(\theta)$ be the conditional probability of z_g, given θ. In the closed response situation, we can write

(6.6)
$$\int_0^1 H_{z_g}(\theta) dz_g = 1 - [P_0(\theta) + P_1(\theta)] \leq 1 ,$$

where $P_0(\theta)$ and $P_1(\theta)$ indicate $P_{z_g}(\theta)$ for $z_g = 0$ and $z_g = 1$, respectively. We can also write for the difficulty parameter b_{z_g}

(6.7)
$$\begin{cases} \lim_{z_g \to 0} b_{z_g} = b_0 \geq -\infty \\ \lim_{z_g \to 1} b_{z_g} = b_1 \leq \infty . \end{cases}$$

We obtain from the definitions of $P^*_{z_g}(\theta)$ and $P_{z_g}(\theta)$

(6.8)
$$P_{z_g}(\theta) \begin{cases} = Q^*_{z_g}(\theta) & z_g = 0 \\ = P^*_{z_g}(\theta) & z_g = 1 , \end{cases}$$

where

$$Q^*_{z_g}(\theta) = 1 - P^*_{z_g}(\theta) \quad . \tag{6.9}$$

It is noted that in the closed response situation inequality holds in (6.6) and in each formula of (6.7) and that we can create each of the other three response situations by setting $P_0(\theta) = 0$ or $P_1(\theta) = 0$, or both.

It is obvious from (6.5) that the operating density characteristic, $H_{z_g}(\theta)$, depends heavily upon the relationship between the item score z_g and the difficulty parameter b_{z_g}, as well as on the functional formula $\Psi_g(.)$. In the closed response situation, the relationship between z_g and b_{z_g} can be any strictly increasing function including the linear function, with the constraint that the values of b_{z_g} are a priori specified at $z_g = 0$ and $z_g = 1$.

For practical purposes, it may be appropriate to consider various strictly increasing polynomials for approximations to such functional relationships. In such approximations, the method of moments for fitting polynomials will be a usefull tool (cf. Samejima and livingston, 1979). We can write for a set of convex polynomials

$$b_{z_g} \doteq b_0 + \sum_{j=1}^{k} \alpha_j \, z_g^j \tag{6.10}$$

with the two constraints,

$$\sum_{j=1}^{k} \alpha_j = b_1 - b_0 \tag{6.11}$$

and

$$\frac{d}{dz_g} b_{z_g} = \sum_{j=1}^{k} \alpha_j \, j \, z_g^{j-1} \geq 0 \qquad 0 < z_g < 1 \quad , \tag{6.12}$$

where strict inequality holds for all values of z_g between zero and unity, except, at most, at an enumerable number of points. A sufficient, though not necessary, condition for the second constraint to hold is that $\alpha_j \geq 0$ for $j = 1, 2, \ldots, k$. We notice that the linear relationship holds by setting $k = 1$. A set of concave polynomials can be obtained under the same condition by

$$b_{z_g} = b_1 - \sum_{j=1}^{k} \alpha_j (1 - z_g)^j \tag{6.13}$$

with the same constraints given by (6.11) and (6.12). Five examples of each of the two sets of polynomials with $k = 1, 3, 5, 7, 9$, and with $\alpha_j = 0$ for $j \leq k$, $b_0 = -2.0$ and $b_0 = 2.0$ are drawn by solid and dotted lines in Figure 6-2, respectively.

In contrast to the observations made so far in the closed response situation, neither in the closed/open response situation nor in the open/closed response situation can the functional relationship between the item score z_g and the difficulty parameter b_{z_g} be linear, nor can it be approximated by a polynomial. One suitable formula in the closed/open response situation may be

$$b_{z_g} = b_0 + \tan[(\pi/2)\xi(z_g)] \tag{6.14}$$

where $\xi(z_g)$ is any strictly increasing, continuous function of z_g defined for $0 \le z_g \le 1$, with the constraint

$$\xi(z_g) \begin{cases} = 0 & z_g = 0 \\ = 1 & z_g = 1 \ . \end{cases} \tag{6.15}$$

Two examples of $\xi(z_g)$ are given by polynomials such that

$$\xi(z_g) = \sum_{j=1}^{k} \alpha_j \, z_g^j \tag{6.16}$$

and

$$\xi(z_g) = 1 - \sum_{j=1}^{k} \alpha_j (1 - z_g)^j \ , \tag{6.17}$$

with the constraints given by the right hand inequality of (6.12) and

$$\sum_{j=1}^{k} \alpha_j = 1 \ . \tag{6.18}$$

Figure 6-3 presents by solid curves five examples of the above functional relationship with (6.16) as $\xi(z_g)$ and with $b_0 = -2.0$, where $k = 1, 3, 5, 7, 9$ and $\alpha_j = 0$ for $j < k$. In the same figure, also presented by dotted curves are the corresponding five examples of (6.14), in which $\xi(z_g)$ is specified by (6.17) with the same set of parameter values.

Figure 6-4 illustrates by a solid curve the operating density characteristic $H_{z_g}(\theta)$ in the normal ogive model as a function of the continuous item score z_g , for each of the four fixed values of θ i.e., -3.0, -2.5, -2.0 and 0.0, with the parameters, $a_g = 1.0$ and $b_0 = -2.0$, using (6.14) as the functional relationship between z_g and b_{z_g} with the specification of $\xi(z_g)$ by (6.16) in which $k = 1$ and $\alpha_1 = 1.0$. In the same figure, also presented by dotted curves are the corresponding operating density characteristics in the logistic model, in which $D = 1.7$.

Similarly, in the open-closed response situation, one useful formula for the relationship between the continuous item score z_g and the difficulty parameter b_{z_g} may be

$$b_{z_g} = b_1 + \tan[(-\pi/2)\varsigma(z_g)] \ , \tag{6.19}$$

where $\varsigma(z_g)$ is any strictly decreasing, continuous function of z_g defined for $0 \le z_g \le 1$, with the constraint

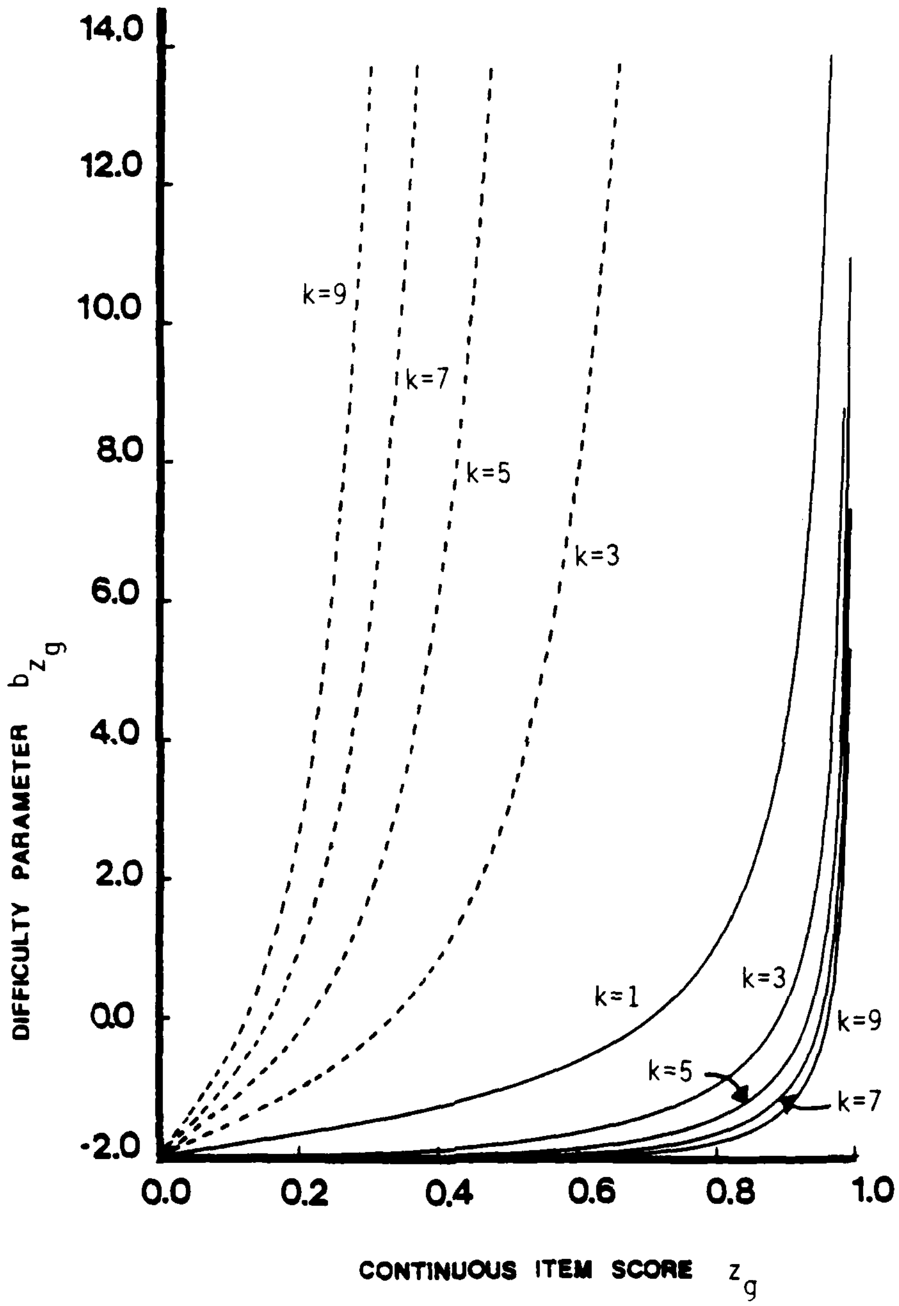


FIGURE 6-3

Five Hypothetical Functional Relationships (Solid Lines) between the Continuous Item Score z_g and the Difficulty Parameter b_{z_g}, Which Are Given by $b_{z_g} = b_0 + \tan[(\pi/2)z_g^k]$ for $k = 1, 3, 5, 7, 9$, with the Parameter, $b_0 = -2.0$, and the Corresponding Relationships (Dotted Lines) Given by $b_{z_g} = b_0 + \tan[(\pi/2)\{1 - (1 - z_g)^k\}]$. Closed/Open Response Situation.

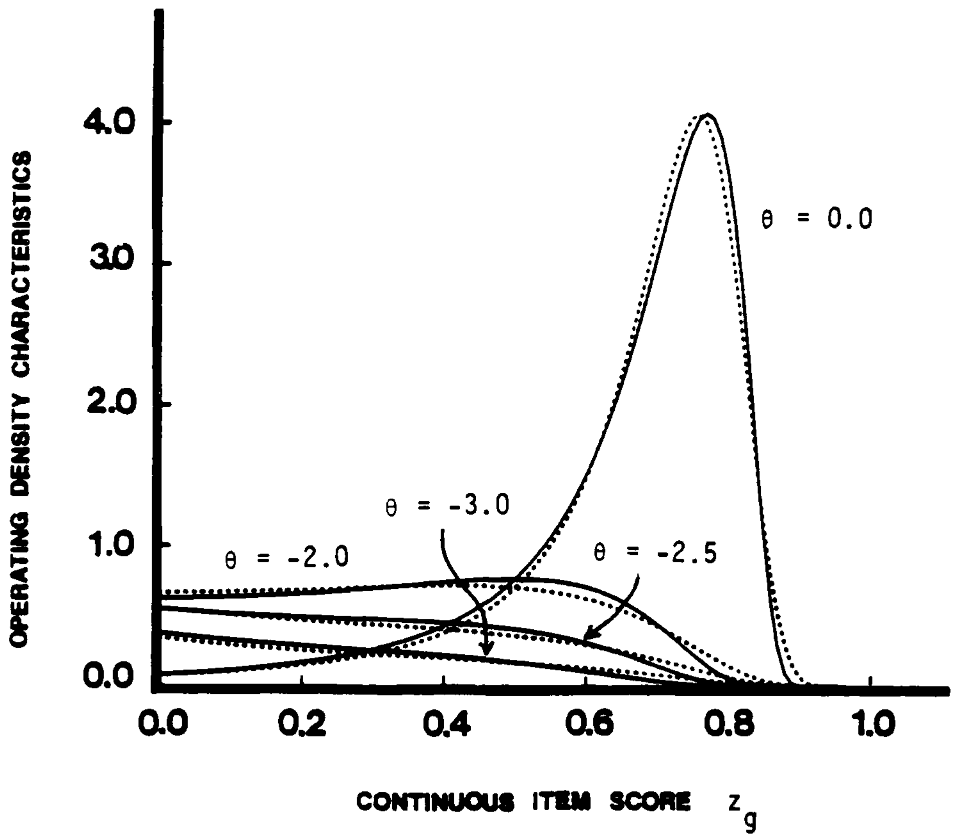


FIGURE 6-4

Operating Density Characteristic, $H_{z_g}(\theta)$, for Each of the Four Values of θ , i.e., -3.0, -2.5, -2.0 and 0.0, Following the Normal Ogive Model with $a_g = 1.0$ and $b_g = -2.0$, with $b_{z_g} = b_0 + [\tan[(\pi/2)z_g]$, Represented by a Solid Curve. Corresponding Four Functions Following the Logistic Model with D = 1.7 Are Also Drawn by Dotted Curves. Closed/Open Response Situation.

$$\varsigma(z_g) \begin{cases} = 1 & z_g = 0 \\ = 0 & z_g = 1 \ . \end{cases} \tag{6.20}$$

Again, for practical purposes, it may suffice if we consider polynomials such that

$$\varsigma(z_g) = \sum_{j=1}^{k} \alpha_j (1 - z_g)^j \tag{6.21}$$

or

$$\varsigma(z_g) = 1 - \sum_{j=1}^{k} \alpha_j \, z_g^j \ , \tag{6.22}$$

where k is the degree of polynomial and α_j $(j = 1, 2, \ldots, k)$ is a coefficient, with the constraints given by the right hand inequality of (6.12) and (6.18). If we set $b_1 = 2.0$ and adopt the parameter values that we used in the examples illustrated in Figure 6-3 for the closed/open response situation, the functional relationships given by (6.19) with (6.21) and (6.22) for $\varsigma(z_g)$ provide us with the set of curves obtained by rotating those of Figure 6-3 by one hundred and eighty degrees, keeping the unit of the ordinate and changing the upper asymptote of the curves to +2.0 . Also by rotating the curves in Figure 6-4 in the three dimensional space as $z_g = 0.5$ as the axis of rotation we obtain the corresponding examples of the operating density characteristics of the open/closed response situation for $\theta =$ 0.0, 2.0, 2.5 and 3.5 , respectively.

[VI.3] Parametric and Nonparametric Estimations of Operating Density Characteristics

In the parametric estimation of the operating density characteristic, some appropriate model must be selected first, and then the estimation is reduced to that of the item parameters that belong to the specific model. Thus, in the parameter estimation, model validation at the end of each stage of research is a necessary and important procedure. We notice that, in the parametric approach, we can always reduce the data based upon the continuous response level to those based either upon the graded response level or upon the dichotomous response level, by categorizing the continuous responses into appropriate discrete response categories. Thus those methods developed for the item parameter estimation on the dichotomous response level (e.g., Lord, 1952, Bock and Aitkin, 1981) and their variations developed for the graded response level, are directly applicable in estimating the item parameters of the operating density characteristics. To be more specific, by adopting an appropriate set of values of z_g , we shall be able to obtain the corresponding set of estimated values of b_{z_g} , and then by an appropriate curve fitting we shall be able to obtain the estimated difficulty parameter function. Since our data on the continuous response level contain more information, in so doing we can also conduct a model validation study, if we design our research appropriately.

In the non-parametric estimation of the operating density characteristics, we assume no mathematical forms a priori. In this direct approach, again we can reduce our data to those which are based upon the graded response level, and those non-parametric methods developed for discrete responses (e.g., Levine, 1980, Samejima, 1977, 1981, and cf. Chapter II) can be applied. If we have *Old Test*, or a set of items whose operating characteristics or operating density characteristics are already known, the application of the techniques will be straightforward. When there is no Old Test, we can select a certain subset of items having high content validity out of all the items in our research, and use this subset in place of the Old Test. In so doing, we may assume several different models for our "Old Test" items, estimate the item parameters using suitable parametric methods, validate or invalidate each model, and select the model which has the highest validity. We may end up with selecting different models for different items. In such a case, as far as each model satisfies the unique maximum condition (Samejima, 1969, 1972, 1973), we can still obtain the maximum likelihood estimate of the subject's latent trait, or individual parameter, by using the basic functions (Samejima, 1969, 1973) based upon those separate models.

In the half-open and half-closed response situations, or in the closed response situation, there is another method, which is a combination of the parametric approach and the non- parametric approach. In the first pair of situations, we can reduce our data to those on the dichotomous response level by using the endpoint with a non-zero probability as one category, and recategorizing all the other continuous responses as the other discrete category. We can use all the items thus dichotomized as the Old Test, searching a suitable model, or models, in the same way described in the preceding paragraph. In the closed response situation, we can trichotomize all the responses, using both endpoints as the lowest and highest categories and all the continuous responses as the intermediate category, and follow the same procedure. We can also conceive of many other variations, depending upon the points where responses are discrete.

The main difference between this new method and the preceding one is that in the new method we make use of all the items used in our research as the Old Test, while in the other only a subset of items is used. In general, the choice of a method should depend upon the nature of our data, including the configuration of the characteristics of our items, the sample size of subjects, and so forth.

[VI.4] Estimation of the Individual Parameters of Subjects

When the item parameters are known, or well estimated, the estimation of the individual parameter, or the point of the latent trait θ at which the subject is located, can be performed through the maximum likelihood estimation. If a simple sufficient statistic for the response pattern V such that

$$V' = (z_1, z_2, \cdots, z_g, \cdots, z_n) \quad . \tag{6.23}$$

does not exist, as is the case with most models, we will use the *basic function* (Samejima, 1969, 1972, 1973) and follow the numerical process to obtain the maximum likelihood estimate $\hat{\theta}_V$ for each response pattern V.

We can write for the general form of the basic function in the closed response situation

$$A_{z_g}(\theta) \begin{cases} = -a_g \, \Psi_g\{a_g(\theta - b_0)\} \, [Q_0^*(\theta)]^{-1} & z_g = 0 \\ = [\frac{\partial}{\partial\theta}\Psi_g\{a_g(\theta - b_{z_g})\}][\Psi_g\{a_g(\theta - b_{z_g})\}]^{-1} & 0 < z_g < 1 \\ = a_g \, \Psi_g\{a_g(\theta - b_1)\} \, [P_1^*(\theta)]^{-1} & z_g = 1 \quad . \end{cases} \tag{6.24}$$

It has been shown (Samejima, 1972) in a somewhat different context that, if $\Psi_g(.)$ follows one of the formulae such as (6.3) and (6.4), those three functions in (6.24) are strictly decreasing in θ , the fact that leads to the unique maximum for the likelihood function $L_V(\theta)$ for each and every response pattern V .

(6.24) also includes the basic functions of all the other three response situations, i.e., they are realized by excluding the line in (6.24) corresponding to each open endpoint. The same rule applies for the item response information function which will be discussed later in this chapter.

In the normal ogive model, which is characterized by (6.3), the basic function takes the form

$$(6.25) \qquad A_{z_g}(\theta) \begin{cases} = -(2\pi)^{-1/2}\, a_g\, exp[-a_g^2(\theta - b_0)^2/2]\, [Q_0^*(\theta)]^{-1} & z_g = 0 \\ = -a_g^2(\theta - b_{z_g}) & 0 < z_g < 1 \\ = (2\pi)^{-1/2}\, a_g\, exp[-a_g^2(\theta - b_1)^2/2]\, [P_1^*(\theta)]^{-1} & z_g = 1 \ . \end{cases}$$

This function is strictly decreasing in θ for all the values of z_g , and, in particular, for $0 < z_g < 1$ it is a linear function with the slope $-a_g^2$ which intercepts the abscissa at $\theta = b_{z_g}$. The two asymptotes of this basic function are zero and negative infinity for $z_g = 0$, positive and negative infinities for $0 < z_g < 1$, and positive infinity and zero for $z_g = 1$, respectively. For the basic function in the logistic model, which is specified by (6.4), we have

$$(6.26) \qquad A_{z_g}(\theta) \begin{cases} = -Da_g\, P_0^*(\theta) & z_g = 0 \\ = Da_g\, [1 - 2P_{z_g}^*(\theta)] & 0 < z_g < 1 \\ = Da_g\, Q_1^*(\theta) & z_g = 1 \ . \end{cases}$$

We can see that this is also strictly decreasing in θ throughout the entire range of θ for each and every item score z_g . It is not a linear function for $0 < z_g < 1$, however, although it also intercepts the abscissa at $\theta = b_{z_g}$. The two asymptotes of the basic function are zero and $-Da_g$ for $z_g = 0$, Da_g and $-Da_g$ for $0 < z_g < 1$, and Da_g and zero for $z_g = 1$, respectively.

The upper graph of Figure 6-5 illustrates five examples of the operating density characteristic, $H_{z_g}(\theta)$, of the continuous item response z_g in the normal ogive model with the item parameters $a_g = 1.0$, $b_0 = -2.0$ and $b_1 = 2.0$, for $z_g = 0.1, 0.3, 0.5, 0.7, 0.9$ in the closed response situation, where the difficulty parameter b_{z_g} is given as the linear function of z_g . In the same graph, also presented by dashed lines are those in the two limiting cases where z_g tends to zero and unity, respectively. The corresponding five operating characteristics and those in the two limiting cases in the logistic model with the same set of item parameters and the scaling factor, $D = 1.7$, are shown in the lower graph of Figure 6-5.

It should be recalled (Samejima, 1973, 1974) that a sufficient statistic, $t(V) = \sum_{z_g \in V} a_g^2\, b_{z_g}$, exists in the normal ogive model in the open response situation. It is not the case with the other three response situations, however, which include $z_g = 0$ or $z_g = 1$, or both, although we can still use $t(V)$ defined

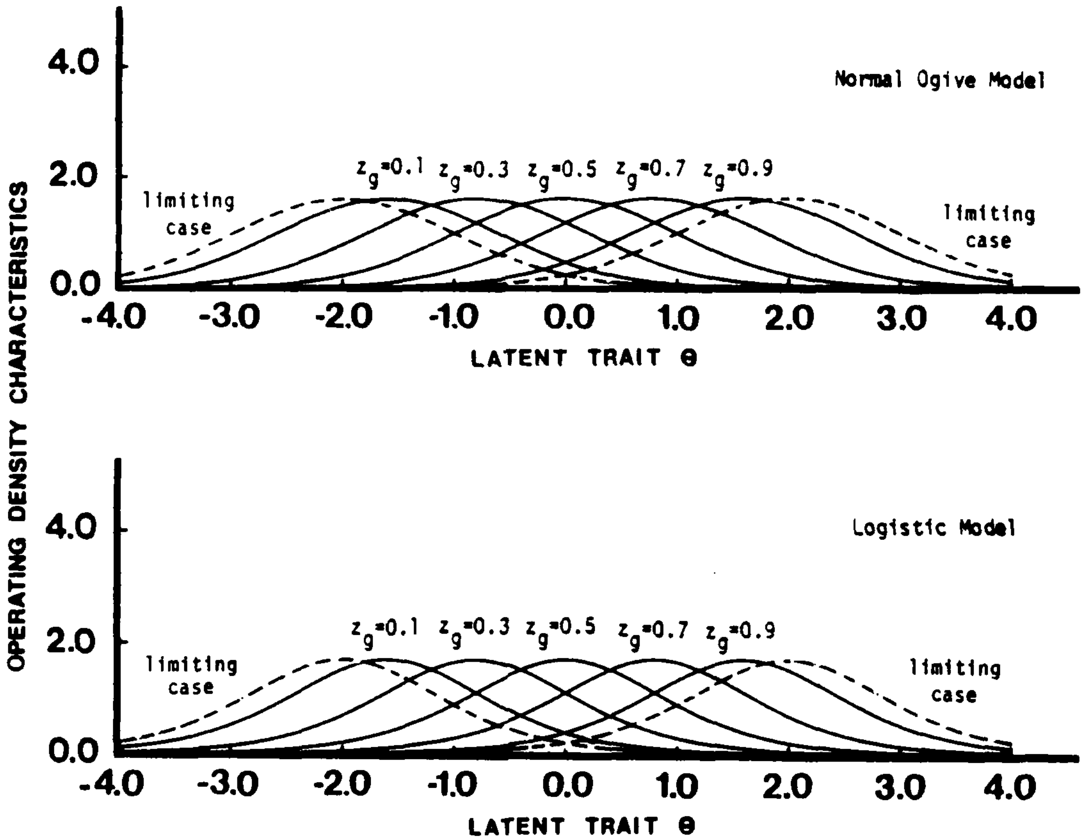


FIGURE 6-5

Operating Density Characteristic, $H_{z_g}(\theta)$, as a Function of θ for Each of the Five Values of the Item Score, 0.1, 0.3, 0.5, 0.7 and 0.9 , Following the Normal Ogive and the Logistic Models, with $a_g = 1.0$, $b_0 = -2.0$, $b_1 = 2.0$ and D = 1.7 , When the Linear Relationship Holds between the Item Score z_g and the Difficulty Parameter b_{z_g} . The Additional Two Curves Are Those in the Limiting Situations Where z_g Tends to Zero and Unity, Respectively. Closed Response Situation.

above for any response pattern which does not include either sero or unity as its elements. It is also recalled (Birnbaum, 1968) that a sufficient statistic, $t^*(V) = \sum_{u_g \in V} a_g u_g$, exists in the logistic model on the dichotomous response level where z_g is replaced by the binary item score u_g . Although the basic functions for $z_g = 0$ and $z_g = 1$ shown in (6.26) are identical with the corresponding functions for $u_g = 0$ and $u_g = 1$ on the dichotomous response level with the replacement of $P_0^*(\theta)$ by $P_1^*(\theta)$, a simple sufficient statistic does not exist, even though $t^*(V)$ can be used for any response pattern which solely consists of 0 and 1 . In general, the maximum likelihood estimation of the individual parameter must be conducted numerically through the basic functions for each response pattern.

For the item response information function, $I_{z_g}(\theta)$, we can write

$$I_{z_g}(\theta) = -\frac{\partial^2}{\partial\theta^2} P_{z_g}(\theta) = -\frac{\partial}{\partial\theta} A_{z_g}(\theta) \tag{6.27}$$

$$\begin{cases} = a_g[(\frac{\partial}{\partial\theta}\Psi_g\{a_g(\theta - b_0)\})Q_0^*(\theta) + a_g(\Psi_g\{a_g(\theta - b_0)\})^2][Q_0^*(\theta)]^{-2} & z_g = 0 \\ = [-(\frac{\partial^2}{\partial\theta^2}\Psi_g\{a_g(\theta - b_{z_g})\})\Psi_g\{a_g(\theta - b_{z_g})\} + (\frac{\partial}{\partial\theta}\Psi_g\{a_g(\theta - b_{z_g})\})^2] & \\ \quad [\Psi_g\{a_g(\theta - b_{z_g})\}]^{-2} & 0 < z_g < 1 \\ = -a_g[(\frac{\partial}{\partial\theta}\Psi_g\{a_g(\theta - b_1)\})P_1^*(\theta) - a_g(\Psi_g\{a_g(\theta - b_1)\})^2][P_1^*(\theta)]^{-2} & z_g = 1\ . \end{cases}$$

In the normal ogive model, this takes the form

$$I_{z_g}(\theta) \begin{cases} = a_g^2\Psi_g\{a_g(\theta - b_0)\}[-a_g(\theta - b_0)Q_0^*(\theta) + \Psi_g\{a_g(\theta - b_0)\}] & \\ \quad [Q_0^*(\theta)]^{-2} & z_g = 0 \\ = a_g^2 & 0 < z_g < 1 \\ = a_g^2\Psi_g\{a_g(\theta - b_1)\}[a_g(\theta - b_1)P_1^*(\theta) + \Psi_g\{a_g(\theta - b_1)\}] & \\ \quad [P_1^*(\theta)]^{-2} & z_g = 1\ , \end{cases} \tag{6.28}$$

and in the logistic model, we obtain

$$I_{z_g}(\theta) \begin{cases} = D^2 a_g^2\ P_0^*(\theta)\ Q_0^*(\theta) & z_g = 0 \\ = 2D^2 a_g^2\ P_{z_g}^*(\theta)\ Q_{z_g}^*(\theta) & 0 < z_g < 1 \\ = D^2 a_g^2\ P_1^*(\theta)\ Q_1^*(\theta) & z_g = 1\ . \end{cases} \tag{6.29}$$

In each of the two models this item response information function is positive throughout the entire range of θ, and the unique maximum condition is satisfied. Figure 6-6 presents the item response information function in the normal ogive model in the upper graph and the one in the logistic model in the lower graph with the same set of parameter values and the scaling factor that we used in Figure 6-5 for $z_g = 0.0,\ 0.1,\ 0.3,\ 0.5,\ 0.7,\ 0.9,\ 1.0$, when the functional relationship between z_g and b_{z_g} is linear. We can see from (6.28) that in the normal ogive model the horizontal line in the upper graph of Figure 6-6 indicates the item response information function for each and every value of z_g in the interval (0,1), so this includes the five cases where $z_g =\ 0.1,\ 0.3,\ 0.5,\ 0.7,\ 0.9$. Those in the two limiting situations where z_g tends to zero and unity, respectively, are also drawn by dashed lines.

The item information function, $I_g(\theta)$, is defined as the conditional mean of the item response information function, given θ (Samejima, 1969, 1972, 1973), for which we can write

$$(6.30) \qquad I_g(\theta) \ = \ I_0(\theta)[1 - P_0^*(\theta)] + \int_0^1 I_{z_g}(\theta) H_{z_g}(\theta) dz_g + I_1(\theta)\ P_1^*(\theta) \ ,$$

where $I_0(\theta)$ and $I_1(\theta)$ indicate the item response information function $I_{z_g}(\theta)$ for $z_g = 0$ and $z_g = 1$, respectively. This function is drawn by a dotted line in each graph of Figure 6-6.

Figure 6-7 illustrates the operating density characteristics $H_{z_g}(\theta)$ in the normal ogive and logistic models in the closed/open response situation, with the item parameters $a_g = 1.0$ and $b_0 = -2.0$, and the scaling factor $D = 1.7$ in the latter model, for the five selected item scores, 0.1, 0.3, 0.5, 0.7, and 0.9 . The difficulty parameter function adopted here is shown in Figure 6-3 as the solid curve marked with $k = 1$. As was observed in the closed response situation, this operating density characteristic is proportional to $\Psi_g(.)$ with a_g^{-1} as the dispersion parameter and b_{z_g} as the location parameter with $a_g(\frac{d}{dz_g} b_{z_g})$ as the ratio of proportionality. Since in this example the derivative of the difficulty parameter function is given by $(\pi/2)\ sec^2[(\pi/2)z_g]$ and it increases with z_g , the area under the curve of $H_{z_g}(\theta)$ in Figure 6-7 increases as z_g does, both in the normal ogive and logistic models. In fact, the area approaches infinity as z_g tends to unity and, therefore, b_{z_g} tends to infinity, the tendency that is hinted by the truncated curves for $H_{z_g}(\theta)$ for $z_g = 0.9$ in the two graphs of Figure 6-7. On the other hand, when the continuous item score z_g tends to zero and, therefore, b_{z_g} tends to b_0 , the ratio of proportionality approaches $(\pi/2)a_g$, and this limiting case of $H_{z_g}(\theta)$ is shown by a dashed curve in Figure 6-7 in each of the normal ogive and the logistic models. The areas under the curves for the same value of z_g across the two graphs of Figure 6-7 are equal.

Figure 6-8 presents the item response information function $I_{z_g}(\theta)$ by solid lines and the item information function $I_g(\theta)$ by a dotted line in each of the normal ogive and the logistic models, with the same parameters, scaling factor, difficulty parameter function and fixed values of z_g as were used in Figure 6-7, together with the limiting case of $I_{z_g}(\theta)$ where z_g tends to zero, which is drawn by a dashed line.

Similar observations were also made for the open/closed response situation, which will not be presented here because of the shortage of space.

[VI.5] Prospect of Adopting These Models for Rorschach Diagnosis

This phase of advancement of latent trait theory dealing with partially continuous and partially discrete responses has enhanced the opportunity of applying the theory for cognitive processes further. For one thing, the model for the closed/open response situation is readily applicable for the response

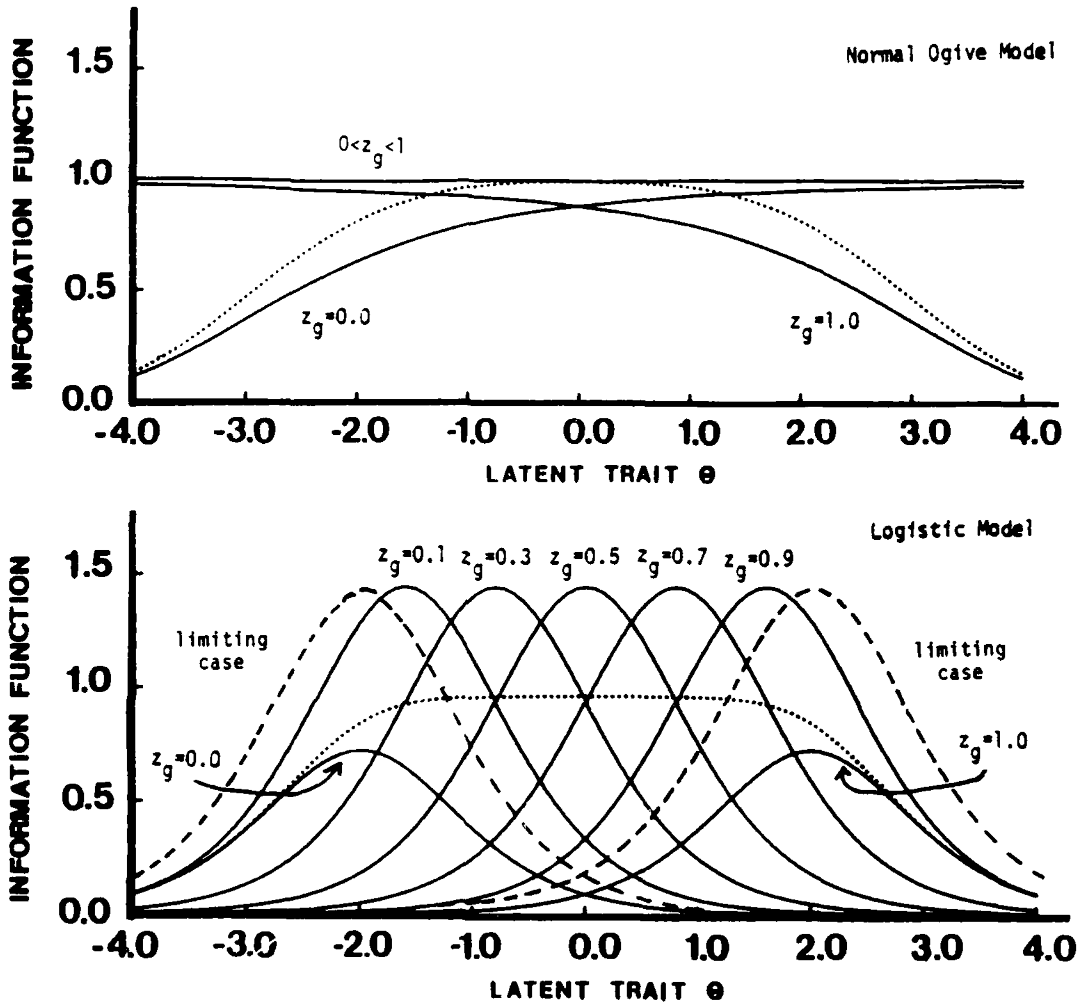


FIGURE 6-6

Item Response Information Functions, $I_{z_g}(\theta)$, (Solid Lines) and Item Information Function, $I_g(\theta)$, (Dotted Line) in the Normal Ogive and the Logistic Models, with $a_g = 1.0$, $b_0 = -2.0$, $b_1 = 2.0$ and D = 1.7 . In the Normal Ogive Model, the Horisontal Line Indicates Common $I_{z_g}(\theta)$ for All Item Scores, $0 < z_g < 1$, While in the Logistic Model the Five Curves Identical in Shape Indicate $I_{z_g}(\theta)$ for $z_g = 0.1, 0.3, 0.5, 0.7, 0.9$, When the Functional Relationship between z_g and b_{z_g} is Linear, with the Two Dashed Curves as Those in the Limiting Situations When z_g Tends to Zero and Unity, Respectively. Closed Response Situation.

latency, which must have a discrete category of "too much delay" or "too slow a response." There are many more conceivable applications of these models, even for kinds of cognitive processes which have never been attempted to be analyzed by psychometric methods.

As one of such ambitious research projects, in 1985 the principal investigator started exploring the possibility of analyzing the clinicians' diagnosis based upon the Rorschach Test. At first she started discussing the idea with one of her colleagues and the director of the Clinical Psychology Program of the University of Tennessee, Dr. Alvin Burstein. Then starting in April, 1986, including a young Ph.D. Scott Glass, the three researchers have met regularly and discussed the new research prospect.

During this period it was decided that we pursue the diagnosis of intellectual aspect of patients as the starter. Although the diagnosis through Rorschach Test is basically for pathological aspects, it is common for clinicians to consider the intellectual aspect of each patient when they decide the therapy, regardless of the specific pathological problem the patient has. In so doing clinicians tend to put more importance upon their own diagnosis through the Rorschach Test than the information given by so-called intelligence tests such as WAIS, WISC, etc. Thus the diagnosis upon the intellectual aspect through the Rorschach Test may be more useful and suitable for us to pursue as the starter, before going into specifics such as schizophrenics, neusotics, etc. Dr. Glass took initiatives in the preliminary study, making various frequency distributions based upon 243 subjects, and also upon randomly selected six subjects out of the more intellectual subgroup, which consists of 68 undergraduate students in the College Scholar Program of the University of Tennessee, and also upon six subjects out of the less intellectual subgroup, i.e., 42 foster care children of twelve to eighteen years of age. Approximately eighty percent of the subjects of this second subgroup have 70 to 80 IQ scores measured by the Peabody Picture Vocabulary Test.

In selecting items, the main task of the principal investigator has been to listen to the two clinicians who are asked to self analyze their ways of diagnosing patients in their intellectual aspect through the Rorschach Test, and also with theoretical considerations to decide the most appropriate way of scoring each item. This has been done repeatedly over the years. Special care has been taken to avoid using overlapping information in defining items and their separate scoring strategies, while taking as much effort as possible to preserve and simulate the actual diagnosis. Recently, Dr. Sandra Loucks also joined our group, and she saw our tentative conclusions in item selection and scoring strategies critically and suggested additions and changes. Also Dr. Allen Rosenwald joined our discussion at one time by our invitation.

Now we have reached the stage that we feel comfortable with our selection of items and their separate scoring strategies, from both the clinicians' and the psychometrician's standpoint. Appendix B presents these results. For each item, the model which is considered most suitable is written, in addition to the content and the scoring strategy of the item. As we can see in this table, these models are the open/closed response model, the closed response model and the graded response model.

This separate research project is still in progress and it will take a long time before we get the results, for we are still in the process of obtaining more Rorschach data and of rescoring each protocol following our definition of items and their separate scoring strategies. The prospect of the success in this research project seems to be good, however, due to the new family of latent trait models proposed in the present research and summarized in this chapter.

[VI.6] Discussion

This proposal of a new set of latent trait models may be one of the biggest accomplishments during this research period. One objective of the proposed research was to bridge psychometrics with cognitive psychology. The principal investigator hopes that in the future these models will be used in different

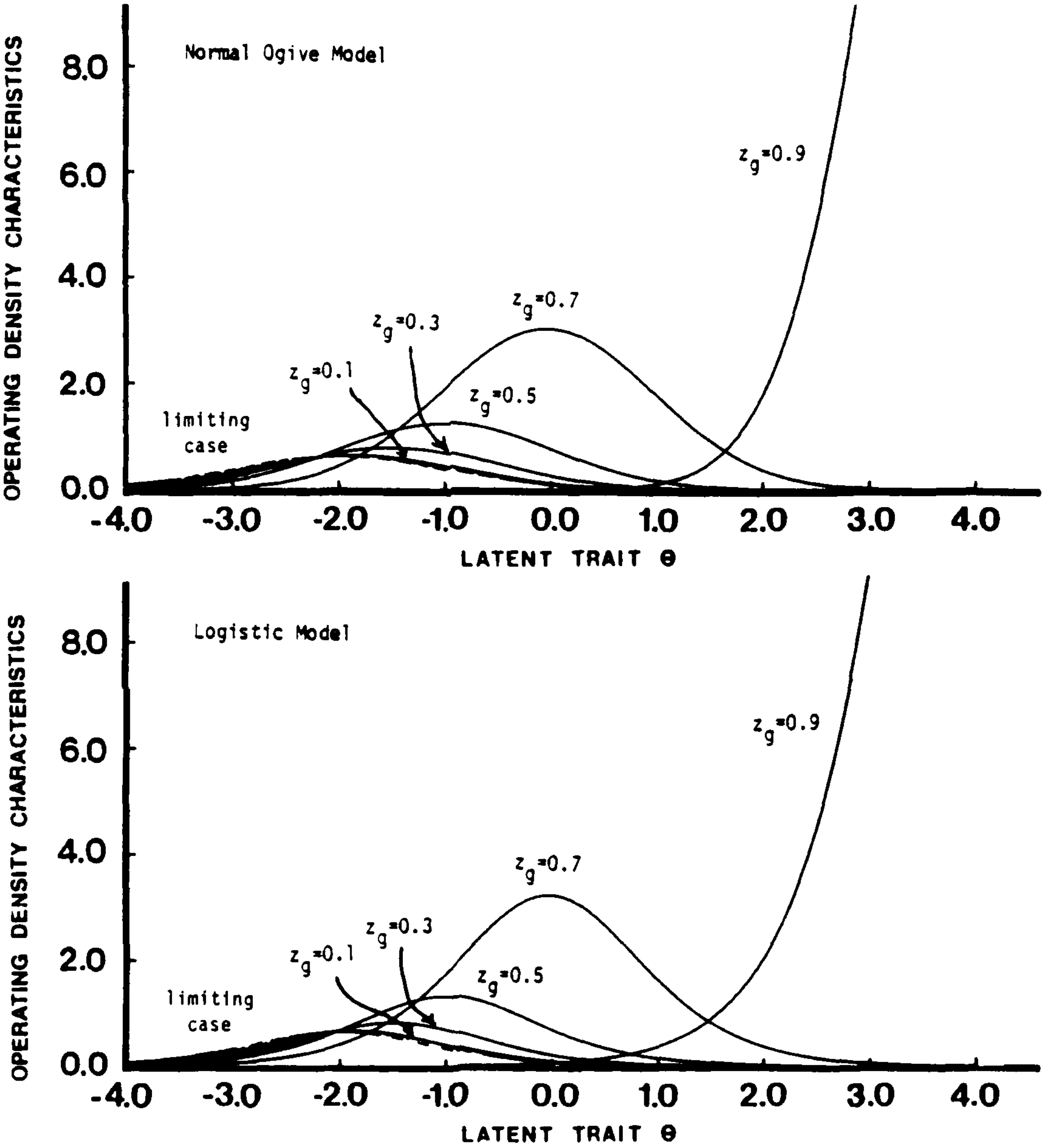


FIGURE 6-7

Operating Density Characteristic, $H_{z_g}(\theta)$, As a Function of θ for Each of the Five Values of the Item Score, 0.1, 0.3, 0.5, 0.7, and 0.9, Following the Normal Ogive and the Logistic Models, with $a_g = 1.0$, $b_g = -2.0$ and D = 1.7 , When the Functional Relationship between the Item Score z_g and the Difficulty Parameter b_{z_g} Is Given by $b_{z_g} = b_0 + \tan[(\pi/2)z_g]$. The Additional Curve Is the One in the Limiting Situation Where z_g Tends to Zero. Closed/Open Response Situation.

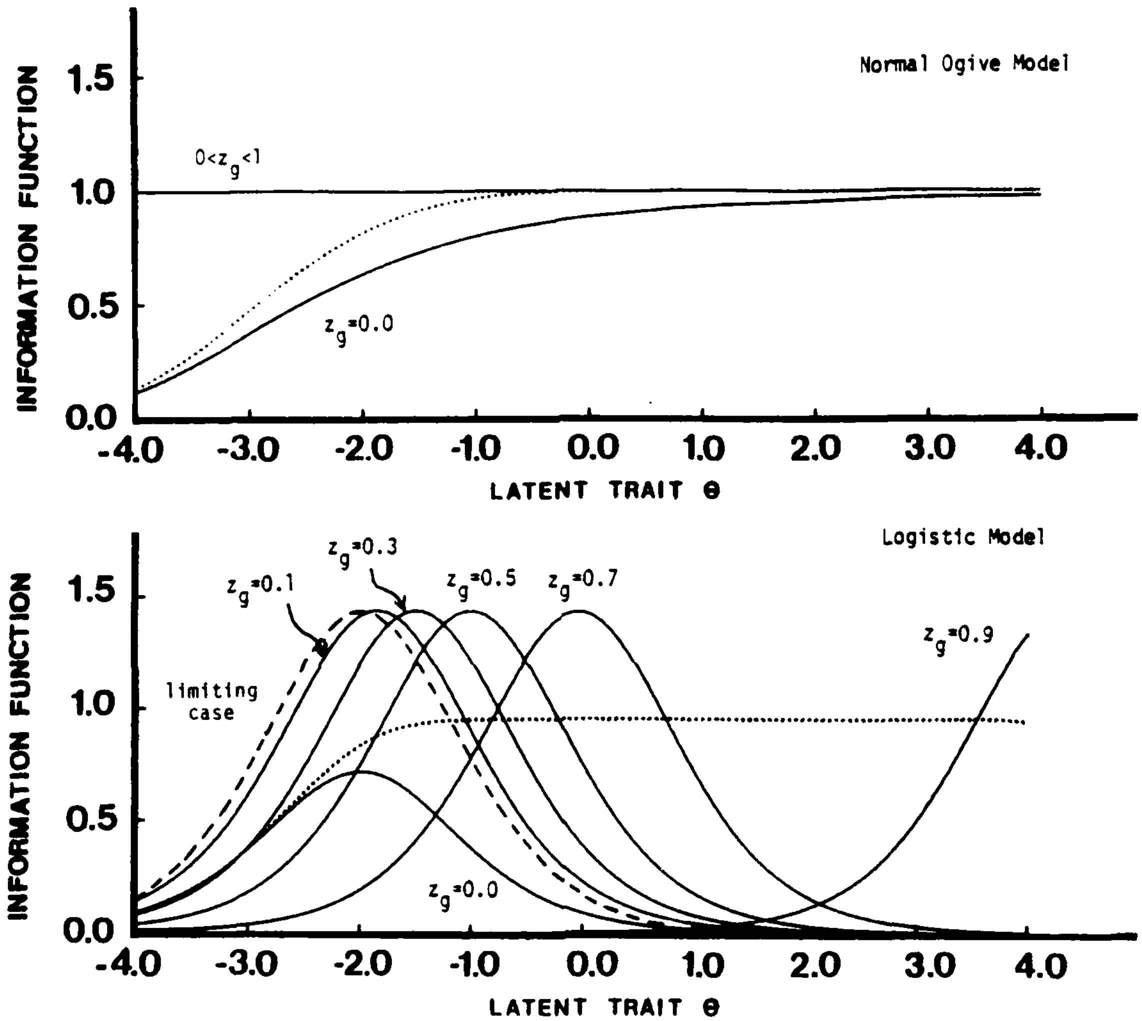


FIGURE 6-8

Item Response Information Functions, $I_{x_g}(\theta)$, (Solid Line) and Item Information Function, $I_g(\theta)$, (Dotted Line) in the Normal Ogive and the Logistic Models, with $a_g = 1.0$, $b_0 = -2.0$ and D = 1.7. In the Normal Ogive Model, the Horizontal Line Indicates Common $I_{x_g}(\theta)$ for All Item Scores, $0 < z_g < 1$, While in the Logistic Model the Five Curves Identical in Shape Indicate $I_{x_g}(\theta)$ for $z_g = 0.1, 0.3, 0.5, 0.7, 0.9$, When the Functional Relationship between z_g and b_{x_g} Is Given by $b_{x_g} = b_0 + \tan[(\pi/2)z_g]$, with the Dashed Curve as the One in the Limiting Situation Where z_g Tends to Zero. Closed/Open Response Situation.

areas of cognitive psychology as well as in those areas where traditionally psychometric theory and methods have been used more frequently.

There are many more graphs which clarify the shapes of various functions developed in this part of research. The reader is directed to [I.1.5] for them.

References

[1] Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, *Statistical theories of mental test scores*. Chapter 17-20. Reading, Mass.: Addison-Wesley, 1968.

[2] Bock, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 1972, 29-51.

[3] Bock, R. D. and Aitkin, M. Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, 46, 1981, 443-459.

[4] Levine, M. Appropriateness measurement and the formula-score method: overview, intercorrelations and interpretations. Paper presented at *the ONR Conference on Model-Based Psychological Measurement*, 1980, Iowa City, Iowa.

[5] Lord, F. M. A theory of test scores. *Psychometric Monograph*, No. 7, 1952.

[6] Samejima, F. Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17, 1969.

[7] Samejima, F. A general model for free-response data. *Psychometrika Monograph*, No. 18, 1972.

[8] Samejima, F. Homogeneous case of the continuous response level. *Psychometrika*, 1973, 38, 203-219.

[9] Samejima, F. Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika*, 1974, 39, 111-121.

[10] Samejima, F. A method of estimating item characteristic functions using the maximum likelihood estimate of ability. *Psychometrika*, 1977, 42, 163-191.

[11] Samejima, F. Final Report: Efficient methods of estimating the operating characteristics of item response categories and challenge to a new model for the multiple-choice item. *Final Report of N00014-77-C-0360, Office of Naval Research*, 1981.

[12] Samejima, F. and Livingston, P. S. Method of moments and the least squares solution for fitting a polynomial. *Office of Naval Research Report*, 79-2, 1979.

VII Informative Distractors and Their Plausibility Functions in the Multiple-Choice Test Items

The multiple-choice format has been most widely used in the paper-and-pencil testing situation, in which the examinee is to choose one of the several alternative answers that are prearranged for each question as his answer. This tendency has also been carried to the computerized adaptive testing, in which the examinee is to answer a sequence of questions selected out of an item pool and presented by a computer.

In dealing with the multiple-choice test item, three-parameter logistic model (Birnbaum, 1968) has been popular among researchers. The model is based upon the knowledge or random guessing principle, which assumes that the examinee either knows the answer or guesses randomly. Thus each item is scored either right or wrong, depending upon whether the examinee has chosen the correct answer or one of the incorrect alternative answers. All incorrect alternative answers are treated as equivalent to each other, therefore, and it is implicitly assumed that those distractors have identical operating characteristics. Such a family of models belongs to the Equivalent Distractor Model.

We notice that Equivalent Distractor Model does not account for any information provided by the choices of specific wrong alternative answers. Thus it treats the multiple-choice test item as nothing but a blurred image of the free-response test item. One must question, however, if indeed the knowledge or random guessing principle fits the examinee's behavior in testing situations. The answer seems to be "No" in most cases, and there exists the examinee's intentional choices of wrong alternative answers, or distractors. Because of this fact, we need some other models that do not belong to the Equivalent Distractor Model.

The principal investigator has proposed a new family of models for the multiple-choice test item (Samejima, 1979), which accounts for such intentional choices of distractors by examinees. In such models, each incorrect alternative answer, as well as the correct answer, provides us with its unique information, although examinees may still guess randomly when they are desperate and have no idea as to which alternative is more plausible than the others as the answer to the question. Such a family of models belongs to the Informative Distractor Model.

The plausibility function of each distractor is defined as the conditional probability assigned to the choice of the particular distractor, given ability. If the plausibility function of one, or more, distractor is informative, then we will be able to make use of the information in ability estimation, as well as the one provided by the correct answer. Thus the multiple-choice test item is no longer a "blurred image" of the free-response test item, but has a unique status as a test item which can be more informative than the free-response test item. The principal investigator's family of models includes these plausibility functions for incorrect alternative answers.

To begin with, it will be worthwhile to estimate the plausibility functions of the distractors of existing test items, in order to find out if, indeed, some distractors provide us with their unique informations. Since we know very little about the behavior of wrong alternative answers of the multiple-choice test item, at this stage it is more desirable to approach their plausibility functions without assuming any mathematical form. Thus, theory and methods for estimating the operating characteristics of discrete item responses, which were summarized in Chapter II, found their full usefulness in this part of research.

In this chapter, a brief outline of the research will be described. For more details and information, see [I.1.6].

[VII.1] Iowa Test Data

Iowa Test Data are based upon the Iowa Tests of Basic Skills, Form 6, Levels 9-14 (Hieronymus and Lindquist, 1971). These tests have been designed, constructed and revised at the College of Education of the University of Iowa since 1935, with the general school population in mind, and basically for the fourth through ninth graders. There are eleven tests in the battery, each of which focuses upon a different basic skill. The numbers of test items in the eleven separate tests vary within the range of 74 through 178, including all the six levels.

Our data were obtained by the courtesy of Professor William Coffman of the University of Iowa. They were collected in three different school systems in the State of Iowa, in the years 1971 through 1977, using the subtests of Levels 11, 12 and 13 (cf. Samejima and Trestman, 1980).

In the present study, the results of the 2,364 examinees on the Level 11 Vocabulary Subtest were most intensively analyzed. This subtest consists of forty-three test items, each of which has four alternative answers, i.e., one correct answer plus three distractors.

[VII.2] Method

We mainly adopted the Simple Sum Procedure of the Conditional P.D.F. Approach combined with the Normal Approach Method (cf. Chapter II) for estimating the plausibility functions. In so doing we needed some suitable substitute for the Old Test, since there is no other set of vocabulary items whose characteristics are already known. In order to handle this situation, we used the Level 11 Vocabulary Subtest itself twice, i.e., first as the Old Test and later as the set of "unknown" test items. Thus on the first stage, each item was scored either "right" or "wrong", and the normal ogive model on the dichotomous response level was assumed. We accepted this model tentatively, and item parameter estimation was performed for each of the forty-three test items. On the second stage, these forty-three test items were treated as "unknown" multiple-choice test items with polychotomous item responses, and for each item we obtained an estimated item characteristic function for the correct answer and an estimated plausibility function for each of the three distractors. The former was then compared with the hypothesized normal ogive function as a part of the model validation process. If the normal ogive model was validated, then we would accept the estimated plausibility functions of the distractors. If not, we would examine the invalidated test items, and either assume more suitable models for them or discard these items, to produce a new Old Test and would repeat the estimation process all over again.

It was assumed that the response tendencies of our 2,364 examinees behind the forty-three test items had a multinormal distribution as their joint distribution. If there existed a single dominating common factor behind these forty-three response tendencies, then it would be defined operationally as the vocabulary ability in question. Consequently, the ability distribution for these 2,364 subjects would also be normal, and the origin and the unit of the scale would be defined at its mean and standard deviation, respectively.

The tetrachoric correlation coefficient was obtained for each pair of test items, using the program written by the principal investigator. The resulting inter-item correlation matrix was factor analyzed, using the computer program for principal factor solution in Biomedical Computer Programs Multivariate Analysis Series 4 (BMDP4M). The communalities were estimated iteratively, with the squared multiple correlation of each variable with all other variables as its initial estimate. If we found a relatively powerful second factor, etc., in addition to the dominating first factor, however, we would eliminate some appropriate items from the Old Test to resolve the clusters, and factor analyze the reduced correlation matrix again, until we reached a single general factor pattern.

The estimated item discrimination parameter, $\hat{a}_g$, and item difficulty parameter, $\hat{b}_g$, were given by

$$(7.1) \qquad \hat{a}_g = \rho_g(1-\rho_g^2)^{-1/2}$$

and

$$(7.2) \qquad \hat{b}_g = \hat{\gamma}_g\, \rho_g^{-1}$$

where ρ_g is the factor loading of item g on the first common factor, and $\hat{\gamma}_g$ is the normal deviate corresponding to the proportion correct p_g of each item g .

[VII.3] Results

The same procedure leading to factor analysis was applied for each of the other ten Level 11 Iowa Subtests, and the resulting sets of eigenvalues are shown in Table 7-1, except for those of the Level 11 Reading Comprehension Subtest (R). We can see that for the Vocabulary Subtest the set of eigenvalues indicates a single common factor structure, although there exist relatively powerful second common factors for several other subtests. This may be due to the fact that reading ability is always required in any subtest in addition to its core performance, while in Vocabulary Subtests those two abilities are close in nature.

As expected, it turned out that the factor loadings on the first common factor were all positive, and, except for those of items 24 and 44 , they are greater than 0.300 , ranging from 0.316 for item 39 to 0.691 for item 30. The largest cluster of factor loadings we can find in those common factors excluding those in the first one is the pair in the fourth factor, i.e., 0.393 for item 33 and 0.368 for item 44. Most of the factor loadings on those other common factors are less than 0.300 in absolute value. From this result, the decision was made to define the first common factor operationally as the vocabulary ability and to use the whole set of items in the Subtest as the Old Test. The estimated item parameters $\hat{a}_g$ and $\hat{b}_g$ are shown in Table 7-2, together with the proportion correct p_g and the fnormal deviate $\hat{\gamma}_g$, for each of the forty-three items.

Figure 7-1 presents the square root of the test information function of the Old Test by a solid line, and also its approximation, i.e., the polynomial of degree 7 obtained by the method of moments using the interval of θ , (-5.0, 5.0), by a dotted line. The actual formula of the latter is given by

$$(7.3) \qquad \begin{aligned} [I(\theta)]^{-1/2} \doteq\ & 3.1915950 - 0.236049720\theta - 0.273225500\theta^2 \\ & + 0.0262482259\theta^3 + 0.0123155780\theta^4 - 0.00114859510\theta^5 \\ & - 0.000227876450\theta^6 + 0.0000183226970\theta^7 \ . \end{aligned}$$

The method of moments was applied for four different intervals of θ , i.e., (-4.0, 4.0), (-4.5, 4.5), (-5.0, 5.0) and (-5.5, 5.5), and the result shown in Figure 7-1 provided us with the best fit. The

TABLE 7-1

Eigenvalues of the Matrix (R-V) for Each of the Ten Level 11 Subtests Obtained As the Results of the Principal Factor Solution of Factor Analysis.

	Tests									
	V	L1	L2	L3	L4	W1	W2	W3	M1	M2
1	11.4174	12.3175	10.5823	11.5618	8.4561	7.8066	6.5457	15.1236	10.2474	7.2963
2	1.0398	1.5332	1.9527	2.1354	1.5431	1.9619	1.0297	4.2759	1.4570	1.4043
3	0.7704	1.0122	1.5139	1.6246	0.9682	0.8137	0.9838	0.9698	0.9571	0.7383
4	0.6788	0.8248	1.0949	1.2001	0.7260	0.7445	0.7495	0.8879	0.7146	0.6322
5	0.6395	0.7283	0.7010	0.9655	0.7014	0.5331	0.6671	0.8028	0.6729	0.5763
6	0.6288	0.6740	0.6257	0.7772	0.5827	0.4974	0.5789	0.7542	0.5537	0.5334
7	0.6023	0.6048	0.6134	0.7368	0.5389	0.4299	0.5468	0.6152	0.5184	0.4784
8	0.5512	0.5207	0.5439	0.5810	0.4632	0.4107	0.4518	0.5983	0.4542	0.4523
9	0.5248	0.4739	0.4916	0.4825	0.4486	0.3541	0.3707	0.5493	0.4138	0.4165
10	0.5084	0.4207	0.4478	0.4174	0.3925	0.3069	0.3143	0.5262	0.4038	0.3271
11	0.4801	0.3885	0.3990	0.3530	0.3589	0.2329	0.2667	0.5190	0.3784	0.2913
12	0.4658	0.3780	0.3545	0.3239	0.3332	0.2237	0.2318	0.4880	0.3544	0.2609
13	0.4471	0.3330	0.3434	0.2952	0.3009	0.1719	0.2151	0.4504	0.3372	0.2399
14	0.4133	0.2928	0.3209	0.2158	0.2859	0.1617	0.1874	0.4165	0.3239	0.2045
15	0.3966	0.2882	0.2981	0.2042	0.2093	0.1427	0.1487	0.3635	0.2803	0.1736
16	0.3725	0.2069	0.2351	0.2022	0.1165	0.1234	0.0849	0.3516	0.2060	0.1218
17	0.3537	0.2026	0.2123	0.1505	0.1101	0.1139	0.0508	0.3268	0.1943	0.0705
18	0.3444	0.1783	0.1903	0.1387	0.0977	0.1005	0.0406	0.3018	0.1692	0.0645
19	0.3188	0.1597	0.1435	0.1329	0.0747	0.0761	0.0236	0.2837	0.1475	0.0384
20	0.3065	0.1351	0.1409	0.1015	0.0638	0.0629	0.0183	0.2500	0.1428	0.0237
21	0.2673	0.1082	0.1127	0.0892	0.0544	0.0441	0.0087	0.2042	0.1339	0.0228
22	0.2574	0.0972	0.0913	0.0702	0.0391	0.0177	-0.0112	0.1980	0.1002	0.0153
23	0.2413	0.0873	0.0893	0.0507	0.0277	0.0070	-0.0228	0.1910	0.0947	-0.0091
24	0.2286	0.0742	0.0706	0.0271	0.0095	-0.0086	-0.0431	0.1778	0.0768	-0.0218
25	0.2161	0.0645	0.0512	0.0176	-0.0077	-0.0324	-0.0758	0.1665	0.0681	-0.0392
26	0.1950	0.0478	0.0332	0.0118	-0.0103	-0.0515	-0.0793	0.1566	0.0517	-0.0452
27	0.1800	0.0381	0.0051	-0.0163	-0.0482	-0.0721		0.1391	0.0331	-0.0594
28	0.1698	0.0236	-0.0066	-0.0312	-0.0656	-0.0955		0.1190	0.0279	-0.0903
29	0.1525	0.0180	-0.0174	-0.0363	-0.0828	-0.1027		0.1059	-0.0025	-0.0991
30	0.1402	0.0060	-0.0329	-0.0475	-0.1108	-0.1164		0.0990	-0.0137	
31	0.1285	-0.0180	-0.0476	-0.0584	-0.1222	-0.1201		0.0837	-0.0306	
32	0.1216	-0.0271	-0.0701	-0.0659	-0.1529	-0.1378		0.0686	-0.0394	
33	0.1139	-0.0385	-0.0747	-0.0851		-0.1500		0.0588	-0.0483	
34	0.0939	-0.0705	-0.1062	-0.0965		-0.1713		0.0437	-0.0808	
35	0.0844	-0.0783	-0.1241	-0.1117		-0.1846		0.0362	-0.0982	
36	0.0605	-0.1021	-0.1373	-0.1298		-0.2370		0.0184	-0.1098	
37	0.0508	-0.1103	-0.1500	-0.1540				0.0151	-0.1244	
38	0.0401	-0.1223	-0.1798	-0.1731				0.0083	-0.1458	
39	0.0172	-0.1322	-0.2007	-0.1923				-0.0138	-0.1598	
40	0.0060	-0.1637	-0.2288	-0.2152				-0.0221	-0.1691	
41	-0.0121	-0.1761						-0.0294	-0.1934	
42	-0.0154	-0.1950						-0.0397	-0.2323	
43	-0.0275	-0.2147						-0.0425		
44								-0.0492		
45								-0.0650		
46								-0.0788		
47								-0.0884		
48								-0.0954		
49								-0.1259		
50								-0.1422		
51								-0.1520		
52								-0.1637		
53								-0.1676		
54								-0.1872		
55								-0.2052		
56								-0.2239		

TABLE 7-2

Estimated Item Discrimination Parameter $\hat{a}_g$ and Item Difficulty Parameter $\hat{b}_g$, Proportion Correct p_g and Normal Deviate $\hat{\gamma}_g$, for Each of the Forty-Three Old Test Items of the Iowa Level 11 Vocabulary Subtest.

Item g	Discrimination Parameter $\hat{a}_g$	Difficulty Parameter $\hat{b}_g$	Proportion Correct p_g	Normal Deviate $\hat{\gamma}_g$
24	0.196	-4.257	0.79315	-0.81740
25	0.829	-1.000	0.73816	-0.63768
26	0.614	-0.821	0.66624	-0.42953
27	0.594	-0.340	0.56895	-0.17370
28	0.669	-0.900	0.69162	-0.50045
29	0.867	-1.077	0.75973	-0.70543
30	0.956	-0.557	0.64975	-0.38465
31	0.938	-0.179	0.54865	-0.12225
32	0.940	-0.803	0.70897	-0.55038
33	0.434	-2.331	0.82318	-0.92755
34	0.598	-1.210	0.73266	-0.62088
35	0.489	-0.569	0.59856	-0.24962
36	0.657	-0.987	0.70601	-0.54177
37	0.351	0.577	0.42428	0.19096
38	0.665	-0.468	0.60237	-0.25949
39	0.333	-0.676	0.58460	-0.21368
40	0.683	0.402	0.41032	0.22672
41	0.531	-0.948	0.67174	-0.44472
42	0.436	0.258	0.45897	0.10303
43	0.672	-0.867	0.68570	-0.48370
44	0.143	4.175	0.27665	0.59282
45	0.898	-0.357	0.59433	-0.23870
46	0.612	-0.318	0.56599	-0.16617
47	0.494	-0.781	0.63536	-0.34608
48	0.849	0.054	0.48604	0.03500
49	0.421	-0.626	0.59602	-0.24306
50	0.346	-0.250	0.53257	-0.08173
51	0.664	-0.420	0.59179	-0.23215
52	0.640	0.217	0.45347	0.11690
53	0.402	0.526	0.42217	0.19635
54	0.573	0.126	0.47504	0.06261
55	0.667	-0.342	0.57530	-0.18988
56	0.593	1.007	0.30372	0.51373
57	0.370	0.398	0.44501	0.13828
58	0.416	0.782	0.38198	0.30028
59	0.491	-0.731	0.62648	-0.32254
60	0.678	-0.170	0.53807	-0.09557
61	0.519	0.748	0.36506	0.34497
62	0.938	-0.485	0.62986	-0.33148
63	0.637	-0.398	0.58460	-0.21368
64	0.818	-0.042	0.51058	-0.02652
65	0.606	0.595	0.37902	0.30806
66	0.604	-0.376	0.57699	-0.19420

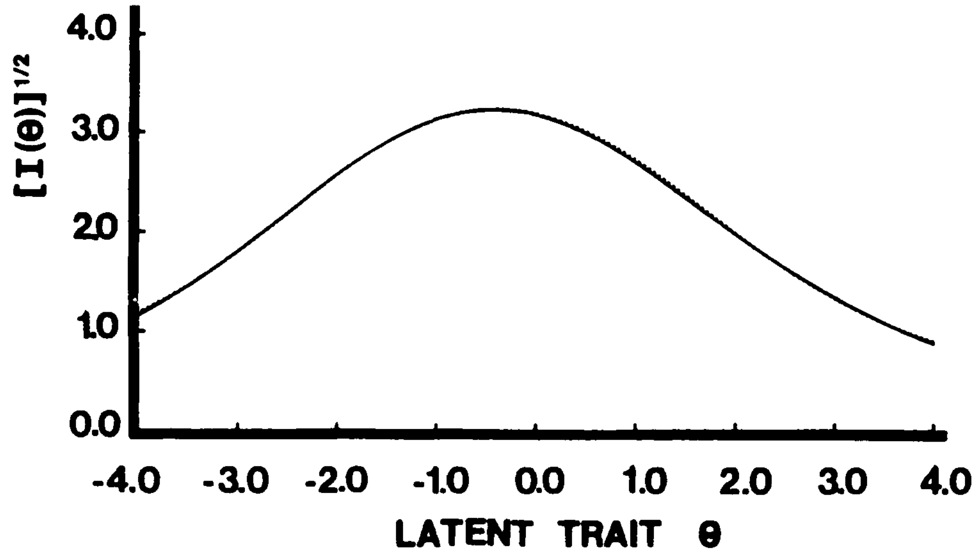


FIGURE 7-1

Square Root of Test Information Function $[I(\theta)]^{1/2}$ of the Level 11 Vocabulary Subtest (Solid Line) and Its Approximation by the Polynomial of Degree 7 Obtained by the Method of Moments with the Specified Interval of θ , [-5.0, 5.0] (Dotted Line).

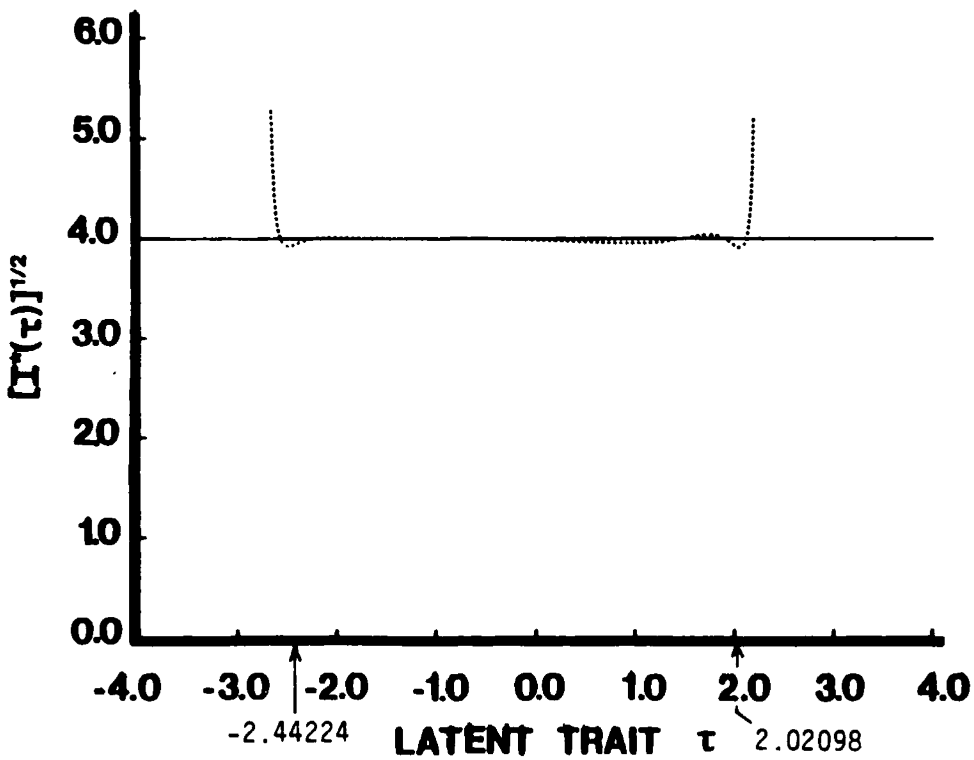


FIGURE 7-2

Square Root of Test Information Function $[I^*(\tau)]^{1/2}$ of the Level 11 Vocabulary Subtest Obtained from the Polynomial Transformation of θ to τ (Dotted Line) and Its Target (Solid Line).

polynomial for transforming θ to τ was obtained from this result, and it turned out to be a polynomial of degree 8 such that

$$(7.4) \qquad \tau(\theta) = 0.00000000 + 0.797898748\theta - 0.0295062156\theta^2 - 0.0227687928\theta^3 + 0.00164051628\theta^4 + 0.000615778918\theta^5 - 0.0000478581278\theta^6 - 0.00000813844468\theta^7 + 0.000000572584288\theta^8 \ .$$

Figure 7-2 presents the square root of the test information function of τ thus obtained by using the approximated polynomial for $[I(\theta)]^{1/2}$ which was given by (7.3) and the derivative of τ obtainable from (7.4). Since the interval of θ, (-4.0, 4.0), corresponds to the interval of τ, (- 2.44244, 2.02098), the latter is shown by arrows in Figure 7-4. We can see that for this interval of τ the approximated square root of the test information function, $[I^*(\tau)]^{1/2}$, is practically constant.

The maximum likelihood estimate, $\hat{\theta}_s$ of θ was obtained for each individual subject from his response pattern on the Old Test items, and was transformed to that of τ through (7.4). On this stage, eight subjects whose $\hat{\theta}_s$ are outside of the interval (- 3.75, 3.75) were excluded permanently from the rest of the research, so that the number of subjects was reduced to 2,356.

The method of moments was applied for the set of 2,356 $\hat{\tau}_s$'s to produce the best fitted polynomials of degrees 3 and 4 in the least square principle (cf. Samejima and Livingston, 1979), and they turned out to be

$$(7.5) \qquad \hat{g}^*(\hat{\tau}) = 0.42358084 - 0.046813019\hat{\tau} - 0.13270786\hat{\tau}^2 + 0.0200142202\hat{\tau}^3$$

and

$$(7.6) \qquad \hat{g}^*(\hat{\tau}) = 0.45023559 - 0.044232853\hat{\tau} - 0.20387563\hat{\tau}^2 + 0.0184068862\hat{\tau}^3 + 0.022176405\hat{\tau}^4 \ ,$$

respectively.

Table 7-3 presents the frequency distribution of the 2,356 $\hat{\tau}'s$ with respect to the types of the conditional distribution of τ, given $\hat{\tau}_s$, in both Degree 3 and 4 Cases. These types, 1 through 7, indicate Pearson's Types (Elderton and Johnson, 1969; Johnson and Kitz, 1970) which were assigned by evaluating the values of the criterion κ. We can see in this table that in both Degree 3 and 4 Cases more than sixty percent of the cases belong to the normal distribution, while most of the others belong to the Beta distribution, i.e., either Pearson's Type 1 or 2. There are some cases whose conditional distributions of τ are undefined, either due to a negative value for an estimated even conditional moment or to a negative value for the estimated conditional probability density. Those subjects were excluded from the rest of the research.

The above results support our choice of the Normal Approach Method in both Degree 3 and 4 Cases. Moreover, a close examination of the skewness and kurtosis indices further discloses the fact that, in

TABLE 7-3

Frequency Distribution of the 2,356 $\hat{\tau}_s$ with Respect to Their Pearson Types for the Conditional Distributions of τ .

Type	Degree 3 Case	Degree 4 Case
1	362	380
2	402	220
3	0	0
4	6	69
5	0	1
6	1	8
7	0	89
normal	1,458	1,536
und. 1	112	47
und. 2	15	6
Total	2,356	2,356

und. 1 : Undefined Due to Negative Even Conditional Moment(s).

und. 2 : Undefined Due to Negative P.D.F.

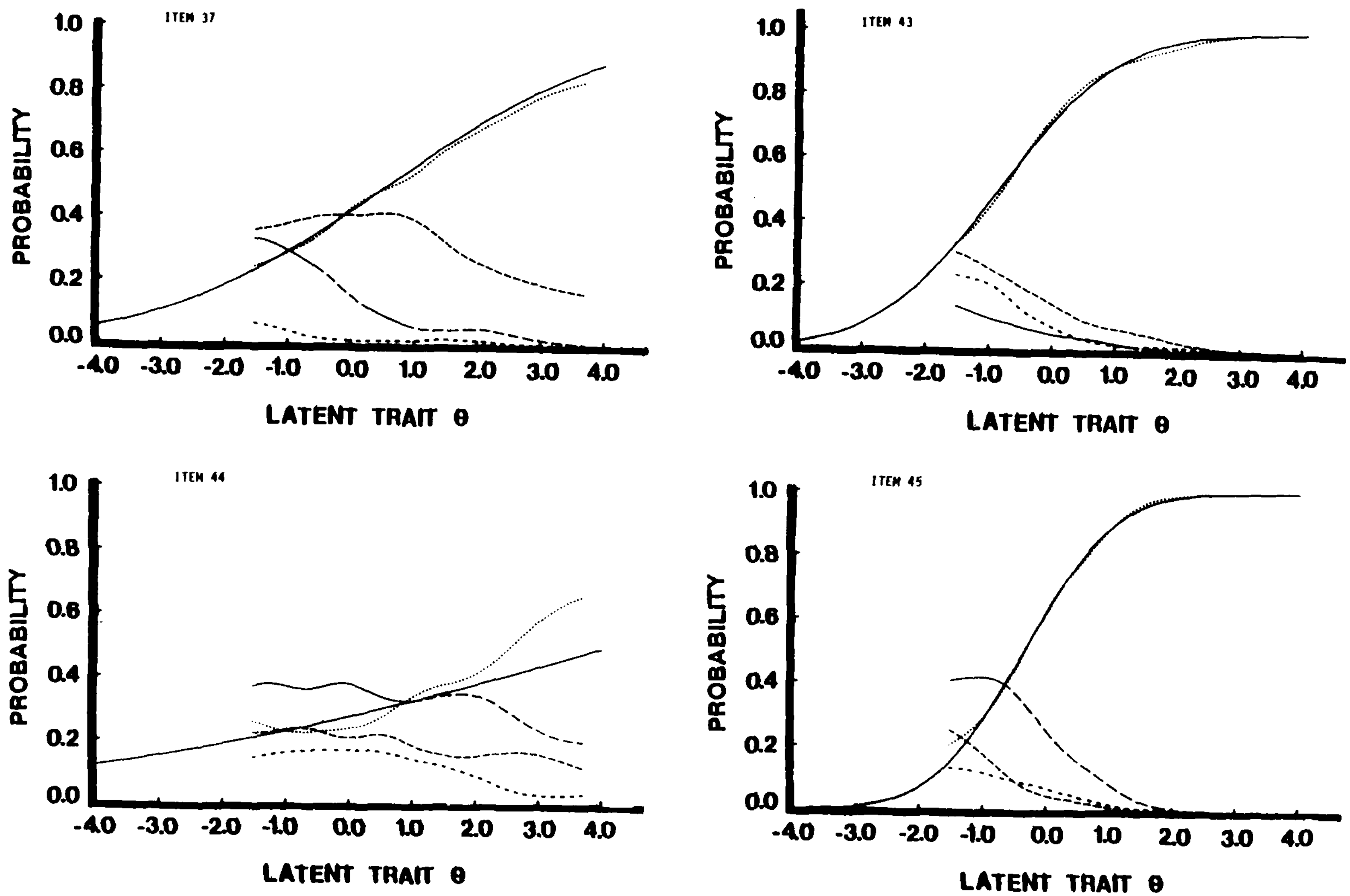


FIGURE 7-3

Estimated Item Characteristic Function (Dotted Line) and Estimated Plausibility Functions (Dashed Lines) for the Three Distractors for Each of the Four Selected Items as Examples.

most cases where the conditional distributions of τ belong to Pearson's Types 1 or 2 distribution, they are very close to 0.0 and 3.0, respectively, i.e., the numbers which characterize the normal distribution.

Since these two sets of results are very similar to each other, from there we dealt solely with Degree 4 Case. It is worth noting, however, that the results of Degree 3 Case would be just as respectable as those of Degree 4 Case, in spite of the fact that the degree of the polynomial approximating $g^*(\hat{\tau})$ is one less and as small as 3.

Figure 7-3 exemplifies the resulting estimated item characteristic function and estimated plausibility functions for each of the four items, i.e., items 37, 43, 44 and 45. For most of the forty-three items, the fitness of the estimated item characteristic function with the initial normal ogive curve, which are drawn by dotted and solid lines, respectively, is as good as that of item 43 or 45, although for some items it is a little worse, as is illustrated in the first graph for item 37. The only exception is item 44, whose four estimated functions are also shown in Figure 7-3. With all these things considered, it was decided to accept the first Old Test, and not to repeat the whole procedure using a modified Old Test. The estimated plausibility functions for items 37 and 45 indicate the existence of some informative distractors for these items, although some other distractors do not explicitly show their informativeness, such as the one drawn by the shortest dashed line in each of the two graphs. For item 43 the three distractors did not prove to be very informative. Note, however, these distractors may be informative on much lower levels of ability θ .

The model validation was further made by computing the chi-square statistics testing the bivariate normality for each pair of response tendencies. The results turned out to be fairly supportive.

[VII.4] Discussion

The item analysis on the Iowa Test Data turned out to be easier and more successful than the principal investigator had anticipated. Most of the test items are not likely to follow the Equivalent Distractor Model, to which the three-parameter logistic or normal ogive model belongs. We have discovered many distractors which are informative, and the results suggest that most of the items follow the Informative Distractor Model. Methodologies involved in the present study appear to be promising, and they will find their usefulness in many other future studies.

The next logical step will be to find out how we can make the best use of the information obtainable from the distractors as well as from the correct answers, in order to increase the efficiency of ability estimation. It is also necessary to collect data for subjects of lower levels of ability in order to find the information provided by all distractors. There is a good prospect that the new family of models for the multiple-choice test item will find its place. In this brief summary only flavor of this part of research was presented. There are many more interesting results for this set of data, and the reader is directed to [I.1.6]. All the results on the other Level 11 Subtests and those on Levels 12 and 13 Subtests are excluded in the present report because of the shortage of space.

References

[1] Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, *Statistical theories of mental test scores.* Chapter 17-20. Reading, Mass: Addison-Wesley, 1968.

[2] Elderton, W. P. and N. L. Johnson. *Systems of frequency curves.* Cambridge University Press, 1969.

[3] Hieronymus, A. N. and E. F. Lindquist. *Iowa Tests of Basic Skills* (Levels Edition), Form 6. Boston: Houghton-Mifflin Company, 1971.

[4] Johnson, N. L. and S. Kotz. *Continuous univariate distributions.* Vol. 2. Houghton Mifflin, 1970.

[5] Samejima, F. Estimation of the operating characteristics of item response categories I: Introduction to the Two-Parameter Beta Method. *Office of Naval Research*, Research Report 77- 1, 1977.

[6] Samejima, F. Estimation of the operating characteristics of item response categories II: Further development of the Two-Parameter Beta Method. *Office of Naval Research*, Research Report 78-1, 1978.

[7] Samejima, F. Estimation of the operating characteristics of item response categories III: The Normal Approach Method and the Pearson System Method. *Office of Naval Research*, Research Report 78-2, 1978.

[8] Samejima, F. Estimation of the operating characteristics of item response categories IV: Comparison of the different methods. *Office of Naval Research*, Research Report 78-3, 1978.

[9] Samejima. F. Estimation of the operating characteristics of item response categories V: Weighted Sum Procedure in the Conditional P.D.F. Approach. *Office of Naval Research*, Research Report 78-4, 1978.

[10] Samejima, F. Estimation of the operating characteristics of item response categories VI: Proportioned Sum Procedure in the Conditional P.D.F. Approach. *Office of Naval Research*, Research Report 78-5, 1978.

[11] Samejima, F. Estimation of the operating characteristics of item response categories VII: Bivariate P.D.F. Approach with Normal Approach Method. *Office of Naval Research*, Research Report 78-6, 1978.

[12] Samejima, F. A new family of models for the multiple-choice item. *Office of Naval Research*, Research Report 79-4, 1979.

[13] Samejima, F. Estimation of the operating characteristics when the test information of the Old Test is not constant I: Rationale. *Office of Naval Research*, Research Report 80-2, 1980.

[14] Samejima, F. Final Report: Efficient methods of estimating the operating characteristics of item response categories and challenge to a new model for the multiple-choice item. *Final Report of N00014-77-C-0360, Office of Naval Research*, 1981.

[15] Samejima, F. and P. S. Livingston. Method of moments as the least squares solution for fitting a polynomial. *Office of Naval Research*, Research Report 79-2, 1979.

[16] Samejima, F. and R. L. Trestman. Analysis of Iowa data I: Initial study and findings. *Office of Naval Research*, Research Report 80-1, 1980.

VIII Analysis of Shiba's Data Collected upon His Word/Phrase Comprehension tests: Comparison of Tetrachoric Method and Logist 5 on Empirical Data

As was pointed out earlier, three-parameter logistic model (Birnbaum, 1968) has earned its popularity in the past years as a model for the multiple-choice test item. This tendency was facilitated by the availability of computer programs, which are represented by Logist 5 (Wingersky, Barton and Lord, 1982), for estimating the three item parameters. Logist 5 can be used not only for the item parameter estimation in the three-parameter logistic model, but also in the (two-parameter) logistic model, by setting the third parameter equal to zero.

In this part of research, Logist 5 and Tetrochoric Method, the latter of which was outlined in the preceding chapter in analyzing the Iowa Test data, were used for the item analysis of empirical data, and comparison was made between the estimated item parameters obtained by assuming the normal ogive model and those obtained by Logist 5 assuming the (two-parameter) logistic model. It is well known (Birnbaum, 1968) that (two-parameter) logistic model provides us with a good approximation to the normal ogive model if we set the scaling factor D equal to 1.7 . In some cases, item parameter estimation was also made by Logist 5 assuming the three-parameter logistic model, and comparison was extended to those results also.

The research report for this part of research contains three hundred thirty-five pages, and it is difficult to even summarize the results. For this reason, only limited illustrations will be given in this chapter. For details and other results and findings, see [I.2.7].

[VIII.1] Shiba's Data

Empirical data adopted here were taken from the test data provided by the courtesy of Professor Sukeyori Shiba of the University of Tokyo, Japan. Shiba's research on the measurement of word/phrase comprehension has been introduced earlier (Samejima, 1980). The battery of tests used for the construction of Shiba's word/phrase comprehension scale consists of thirteen tests, AP1, AP2, A1, A2, A3, A4, A5, A6, J1, J2, S1, S2 and U . Each of these thirteen tests contains thirty to sixty multiple-choice items, each of which has a set of five alternatives. These tests differ in difficulty, and each is designed for a different age group of subjects, ranging from four years of age to the ages of college students. There are subsets of items included in two tests, which are adjacent to each other in difficulty. For example, items 37 through 56 of Test J1 are also items 1 through 20 of Test J2. There are 480 test items in total. The number of examinees used by Shiba for the ability scale construction varies between 219 preschoolers for Test AP1 and 924 second graders of senior high schools for Test S1 .

The principal investigator has found Shiba's tests very well constructed. Professor Shiba and she have been collaborating for the past ten years, and she decided to adopt Shiba's data for this part of research. Some of the results obtained by the Tetrochoric Method were taken from Professor Shiba's work itself.

Out of Shiba's thirteen tests, four tests were chosen for the present research, i.e., A5, A6, J1 and J2. The examinees who took these four tests in Shiba's original data are as the following.

Test A5: 599 fifth graders in elementary schools
Test A6: 412 sixth graders in elementary schools

Test J1: 614 first graders in junior high schools
Test J2: 758 third graders in junior high schools

These groups of examinees and their performances are called, for brevity, A5/0599 Case, A6/0412 Case, J1/0614 Case and J2/0758 Case, respectively. There are also 461 second graders in junior high school who took Test J1 in Shiba's original data. In order to increase the number of examinees, this group of 461 subjects and their performances were added to the J1/0614 Case, to provide us with the J1/1075 Case. This case was further joined by an additional group of 1,184 students of four different junior high schools in Tokyo, to whom Test J1 was administered in some other research of Shiba's. We shall call this largest group of examinees and their performances J1/2259 Case. Thus we have six cases in total, with three of them partly overlapping.

When the item parameter estimation was made by Logist 5, in some cases two or more tests and the corresponding samples of examinees were combined, in order to increase the number of test items and hence to improve the accuracy of estimation. Table 8-1 presents the resulting combinations of tests and the numbers of examinees. When two or more adjacent tests are combined, the number of items is less than the sum total of the numbers of items of the separate tests because of the overlapping items.

[VIII.2] Results

Figure 8-1 presents the estimated ability distribution of each of the original and combined examinee groups, which was obtained through the item parameters estimated by the Tetrochoric Method.

Figure 8-2 shows eight scatter diagrams of the estimated item discrimination parameters of Test J1. They consist of four pairs, in each of which the Logist 5 results of "c_g-zero" (left) and "c_g-free" (right) are compared with the results of the Tetrachoric Method. These results are obtained for different cases, and they are specified in the captions of separate pairs. We can see a substantial consistency between the two sets of estimated item discrimination parameters in the first graph of each of the four pairs, i.e., when (two-parameter) logistic model is assumed in using Logist 5 , whereas there exists little consistency in the second graph of each pair, i.e., when three-parameter logistic model is assumed. We notice that the greatest consistency is observed in the first graph of the first pair of scatter diagrams and in the first pair of the third pair. They are Case J1/1075:c_g-zero of Logist 5 against J1/1075 Case of the Tetrachoric Method and Case J1/2259:c_g-zero against J1/2259 Case of the Tetrachoric Method, i.e., the only two situations which concern the same examinee group both in using Logist 5 and in using Tetrachoric Method, and no guessing parameter is assumed in using Logist 5 . This fact suggests that these two methods provide us with consistent results when the item parameter configurations are such as those of Test J1, if the sample size is 1,000 or above. The corresponding eight scatter diagrams for the estimated item difficulty parameters are presented in Figure 8-3. We can see a similar tendency as we have observed for the estimated discrimination parameters, although inconsistency between the two sets of estimates is less conspicuous when three-parameter logistic model is assumed in using Logist 5. These tendencies were carried out almost as they are even after certain scale adjustments had been made in order to make the comparison more adequate (cf. [I.2.7], Chapters 6 and 8).

Figure 8-4 presents four graphs which clarify how estimated item parameters differ when three-parameter logistic model is assumed in comparison with those when (two-parameter) logistic model is assumed. The first two graphs concern with the group of 1,075 examinees and the last two with the group of 2,259 examinees, and in each pair the first graph concerns with the adoption of the (two-parameter) logistic model and the second with the three-parameter logistic model. In each graph, the estimated item difficulty parameters of the items of Test J1 are taken on the abscissa, and the estimated

TABLE 8-1

Tests, Numbers of Items, Number of Examinees and Other Information for Thirteen Different Cases.

Method		Test(s)	No. of Examinees	Original No. of Items	Excluded Items	No. of Items Included
Tetrachoric Method (Shiba)		A5	599	48	3,13,17	45
		A6	412	56	--	56
		J1	614	56	38	55
		J2	758	60	2	59
Tetrachoric Method (Samejima)		J1	1,074	56	38	55
		J1	2,259	56	38	55
Logist 5	0.0	A5,A6	1,011	88	--	88
	0.0	J1,J2	1,833	96	--	96
c_g:	0.0	A5,A6,J1,J2	2,844	168	--	168
	0.0	J1	1,075	56	38	55
	free	J1	1,075	56	38	55
	0.0	J1	2,259	56	38	55
	free	J1	2,259	56	38	55

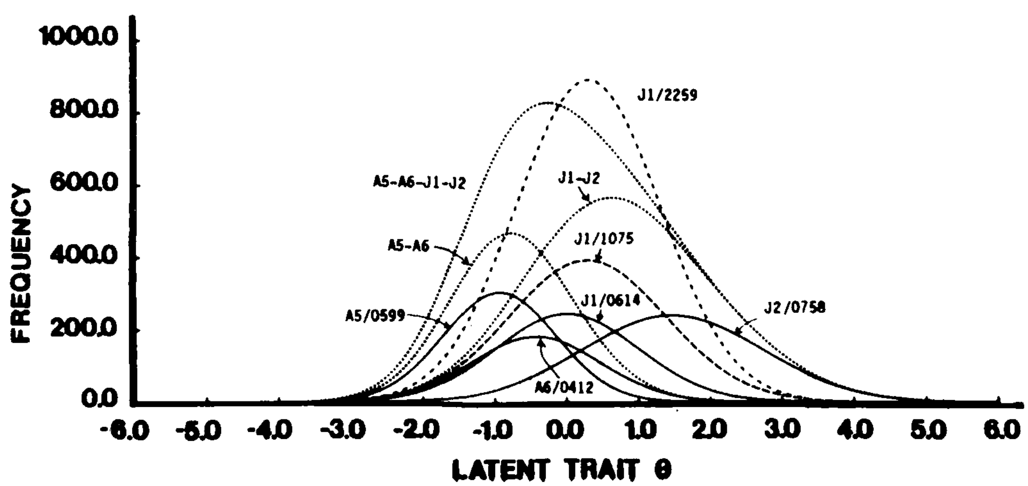


FIGURE 8-1

Estimated Ability Distributions (Solid Lines) of the A5/0599, A6/0412, J1/0614 and J2/0758 Cases, Those of the J1/1075 Case (Long Dashed Line) and of the J1/2259 Case (Short Dashed Line), Together with Those (Dotted Lines) of the Combined Examinee Groups, A5-A6 , J1-J2 and A5-A6-J1-J2.

discrimination parameters are taken on the ordinate. Both the results of the Tetrachoric Method and those of Logist 5 are plotted in each of the four graphs, to make the total number of points 110. To avoid confusion, there are five different symbols, i.e., ▲ , ✱ , ♠ , ◆ and ✚ , in these graphs, and an arrow is drawn for each item from the point indicating the result of the Tetrachoric Method to that of Logist 5 .

Comparison of the first graph of Figure 8-4 with the second, and of the third graph with the fourth, discloses how radically the two estimated parameters of these items of Test J1 are enhanced because of the existence of the guessing parameter c_g when, in using Logist 5, three-parameter logistic model is assumed. These tendencies are similarly observed in both pairs, where the examinee groups of 1,075 individuals and of 2,259 examinees are involved, respectively. These tendencies were carried out almost as they are even after a certain scale adjustment had been made in order to make the comparison more adequate (cf. [I.2.7], Chapter 8).

Table 8-2 presents the direct estimates of the mean and the standard deviation of the distribution of the maximum likelihood estimate of ability, $\hat{\theta}$, for each of the five examinee groups, which are shown as Examinee Group 2, obtained from the combined linear relationship in each case. Since there are more than one way of obtaining these two values in each case, all of these results are presented in Table 8-2. In this table, Examinee Group 1 indicates the group of examinees upon which the item parameters were estimated by the Tetrachoric Method, and Examinee Group 2 means the group of individuals upon which they were estimated by Logist 5. Thus the mean and the standard deviation of the distribution of $\hat{\theta}$ of Case J1-J2, for example, can be estimated in two ways, i.e., through the scatter diagrams of the estimated parameters of the items of Test J1, and through those of the items of Test J2. In the same table, the weighted averages of the estimated mean and standard deviation of the distribution of $\hat{\theta}$ are also given for each examinee group. The weight adopted here is the number of the examinees in Examinee Group 1. We can see in these results that most estimates for the same group of examinees are close to each other.

Figures 8-5 through 8-8 present five estimated item characteristic functions for each of four items of Tests A5, A6, J1 and J2, respectively. In each graph, the result based upon the estimated item parameters obtained by the Tetrachoric Method is drawn by a solid line, and all the other four curves of different lengths of dashes concern with those based upon the estimated item parameters by Logist 5 . We notice that there are basically two sets of estimated item parameters which were obtained by Logist 5, i.e., one based upon either Case A5-A6 or Case J1-J2 and the other upon Case A5-A6-J1-J2. For brevity, we shall call the former approach Method A and the latter Method B. In each of these two cases the estimated item parameters were adjusted twice, i.e., first on the assumption that the mean and the standard deviation of the distribution of $\hat{\theta}$ are the same as those of the distribution of θ , and, secondly, without this assumption. In each graph of Figures 8-5 through 8-8, the results based upon Method A and upon the first and the second scale adjustments are drawn by a long dashed line and a short dashed line, respectively, and those based upon Method B and upon the first and the second adjustments are shown by a dashed line of medium length and dotted line, respectively.

From these results, we can say the following.

(1) For many items, the two Logist 5 results based upon the second scale adjustment are close to each other, while those based upon the first scale adjustment are substantially different from each other.

(2) In addition to the above, the two Logist 5 results based upon the second scale adjustment also tend to be closer to the result of the Tetrachoric Method.

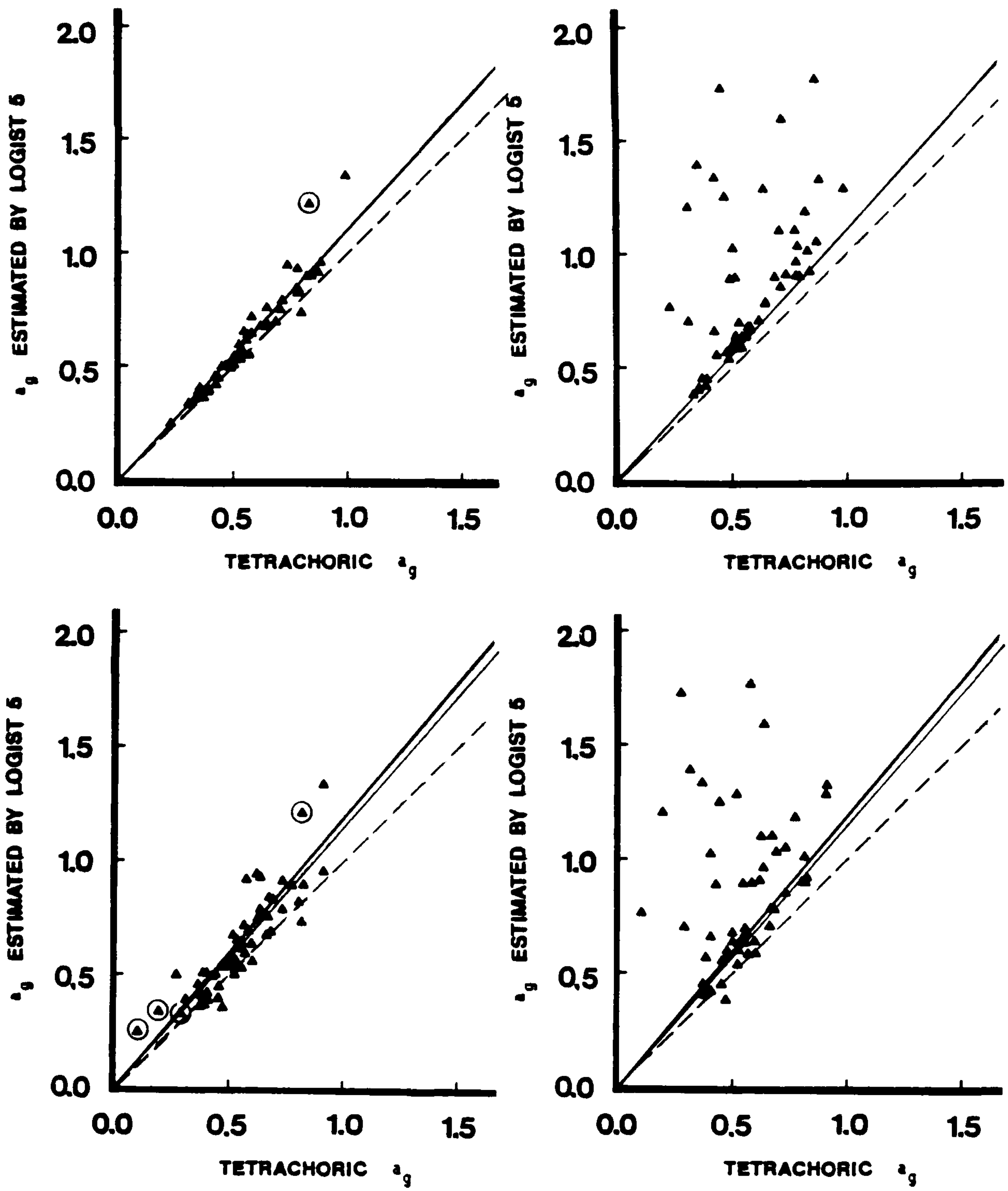


FIGURE 8-2

Estimated Item Discrimination Parameters of the 55 Items of Test J1 Obtained by Logist 5 Plotted against Those Obtained by the Tetrachoric Method. In Using Logist 5, Logistic Model Is Assumed in the Graph on the Left Hand Side and Three-parameter Logistic Model Is Assumed in the Graph on the Right Hand Side. Both Sets of Estimates in Each Graph are the Original One, i.e. before Any Scale Adjustment. The Best Fitted Linear Relationship When the Logistic Model is Assumed is Drawn by a Thin, Solid Line and the One Based upon Both Parameters Are Shown by a Thick, Solid Line. J1/1075 Case (Upper Graphs) and J1/0614 Case on the Abscissa and J1/1075 Case on the Ordinate (Lower Graphs).

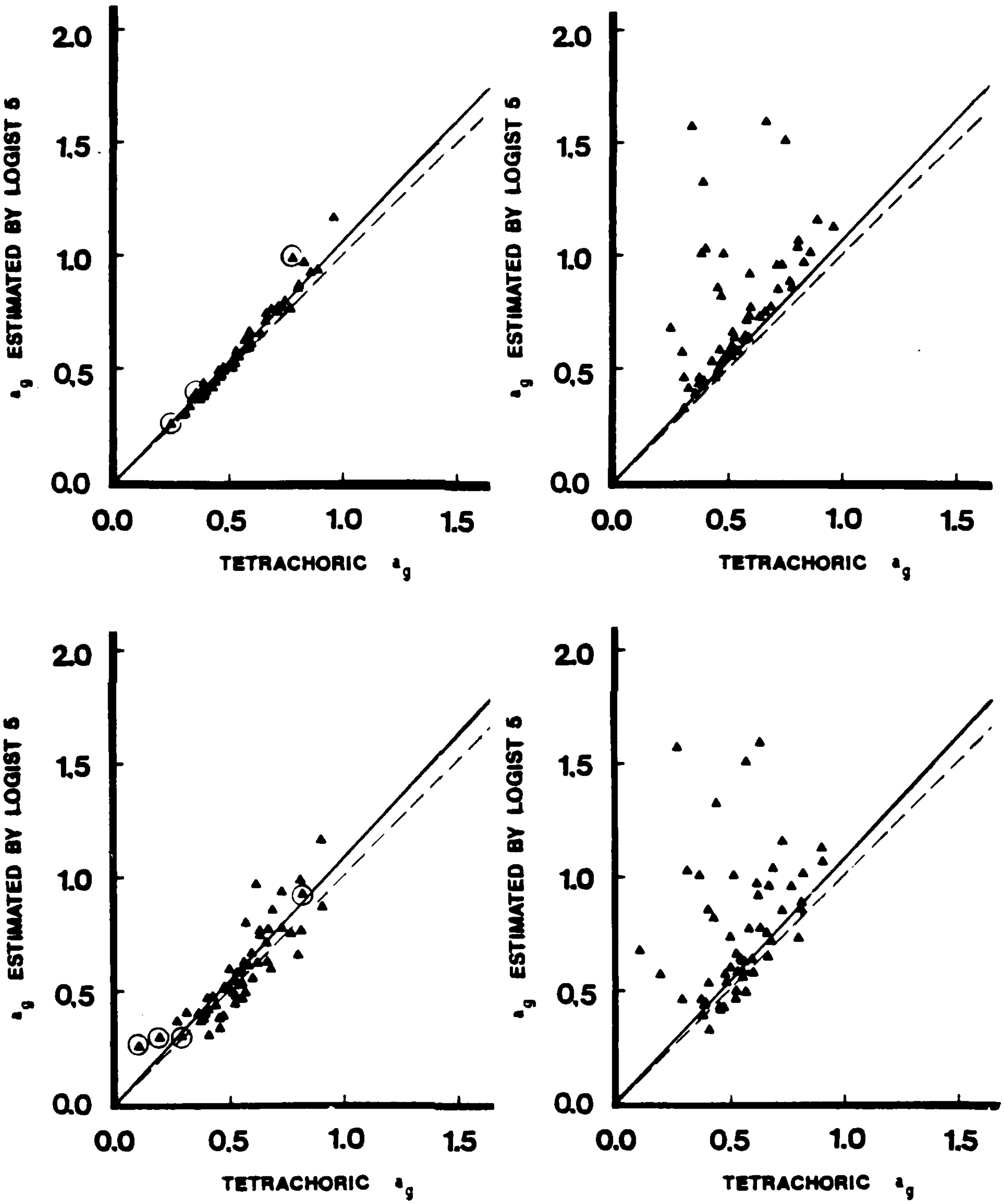


FIGURE 8-2 (Continued)

J1/2259 Case (Upper Graphs) and J1/0614 Case on the Abscissa and J1/2259 Case on the Ordinate (Lower Graphs).

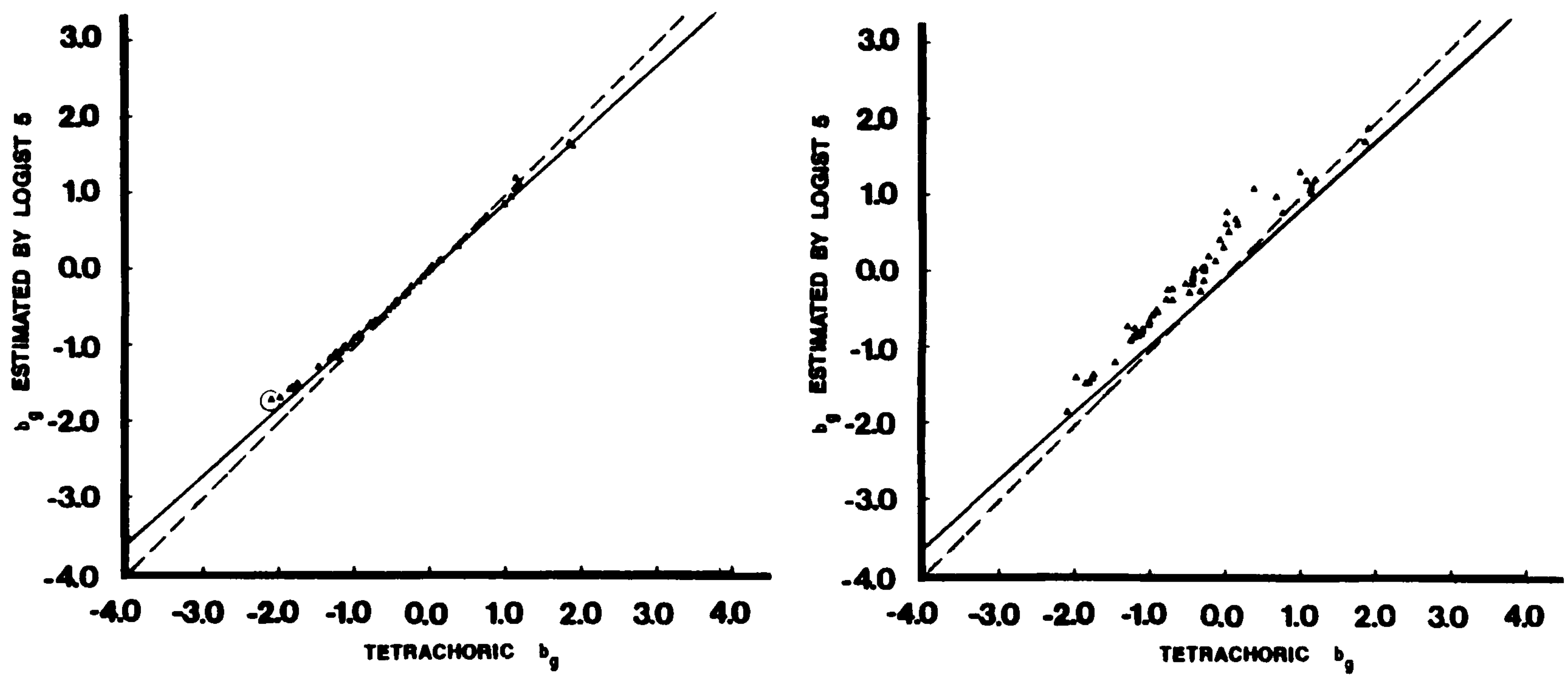


FIGURE 8-3

Estimated Item Difficulty Parameters of the 55 Items of Test J1 Obtained by Logist 5 Plotted against Those Obtained by the Tetrachoric Method. Both Sets of Estimates in Each Graph Are the Original One, i.e., before Any Scale Adjustment. The Best Fitted Linear Relationship When the Logistic Model is Assumed is Drawn by a Thin, Solid Line and the One Based upon Both Parameters Are Shown by A Thick, Solid Line. Logistic Model (Left Graph) and Three-Parameter Logistic Model (Right Graph) Are Assumed in Using Logist 5, Respectively. J1/1075 Case.

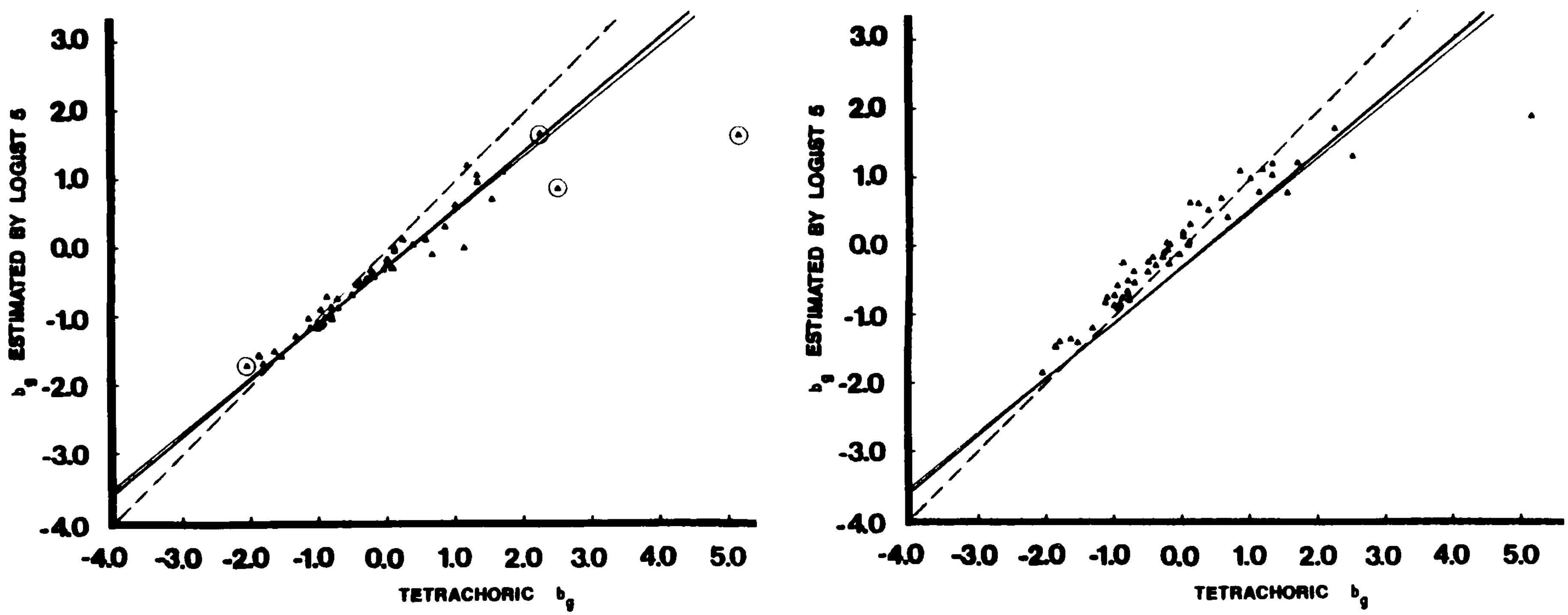


FIGURE 8-3 (Continued)

J1/0614 Case on the Abscissa and J1/1075 Case on the Ordinate.

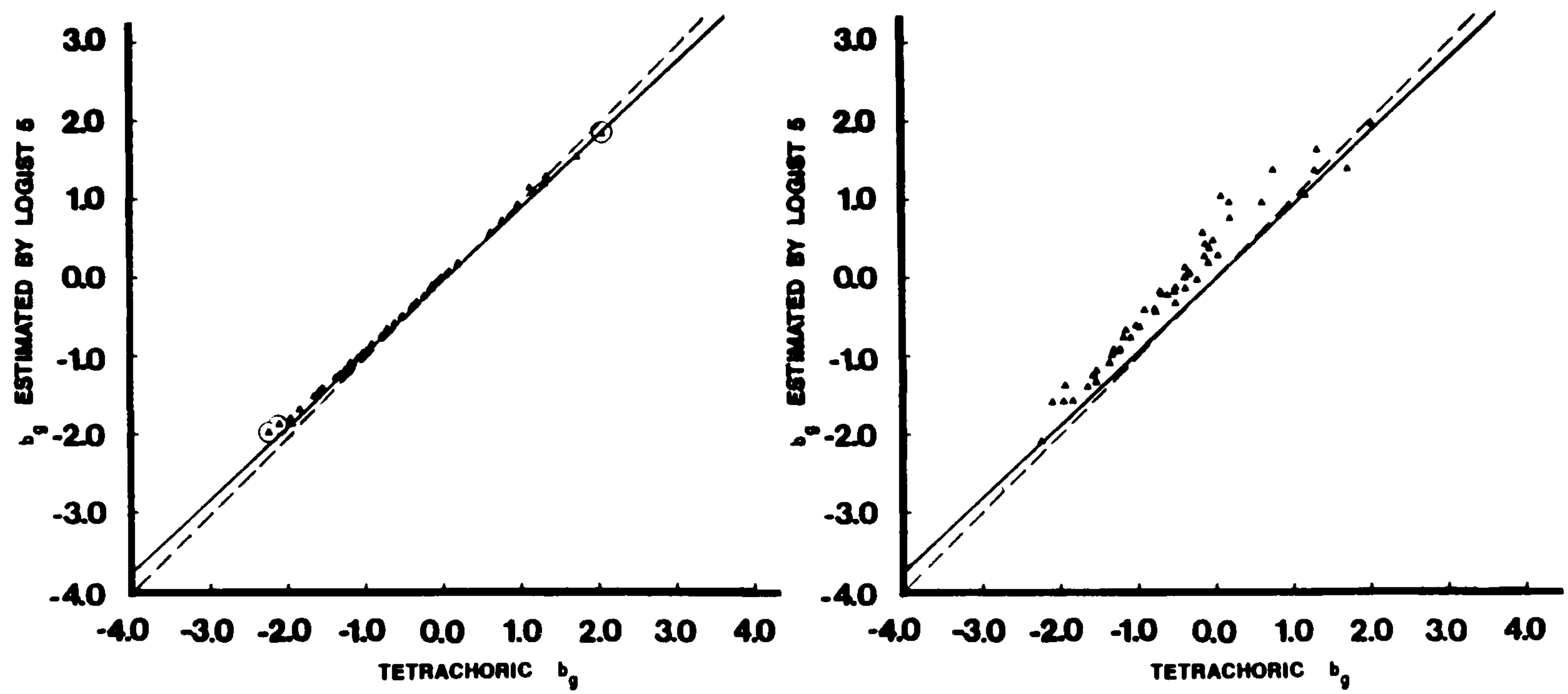


FIGURE 8-3 (Continued)

J1/2259 Case.

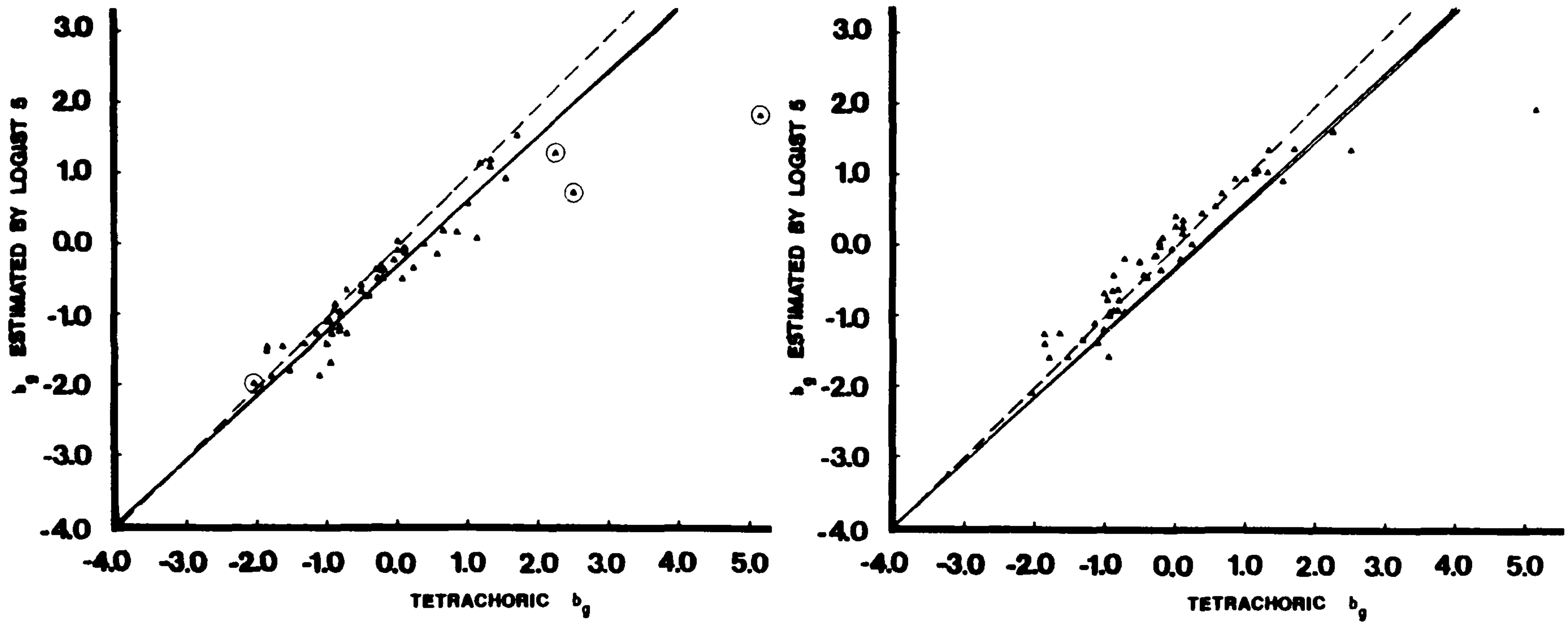


FIGURE 8-3 (Continued)

J1/0614 Case on the Abscissa and J1/2259 Case on the Ordinate.

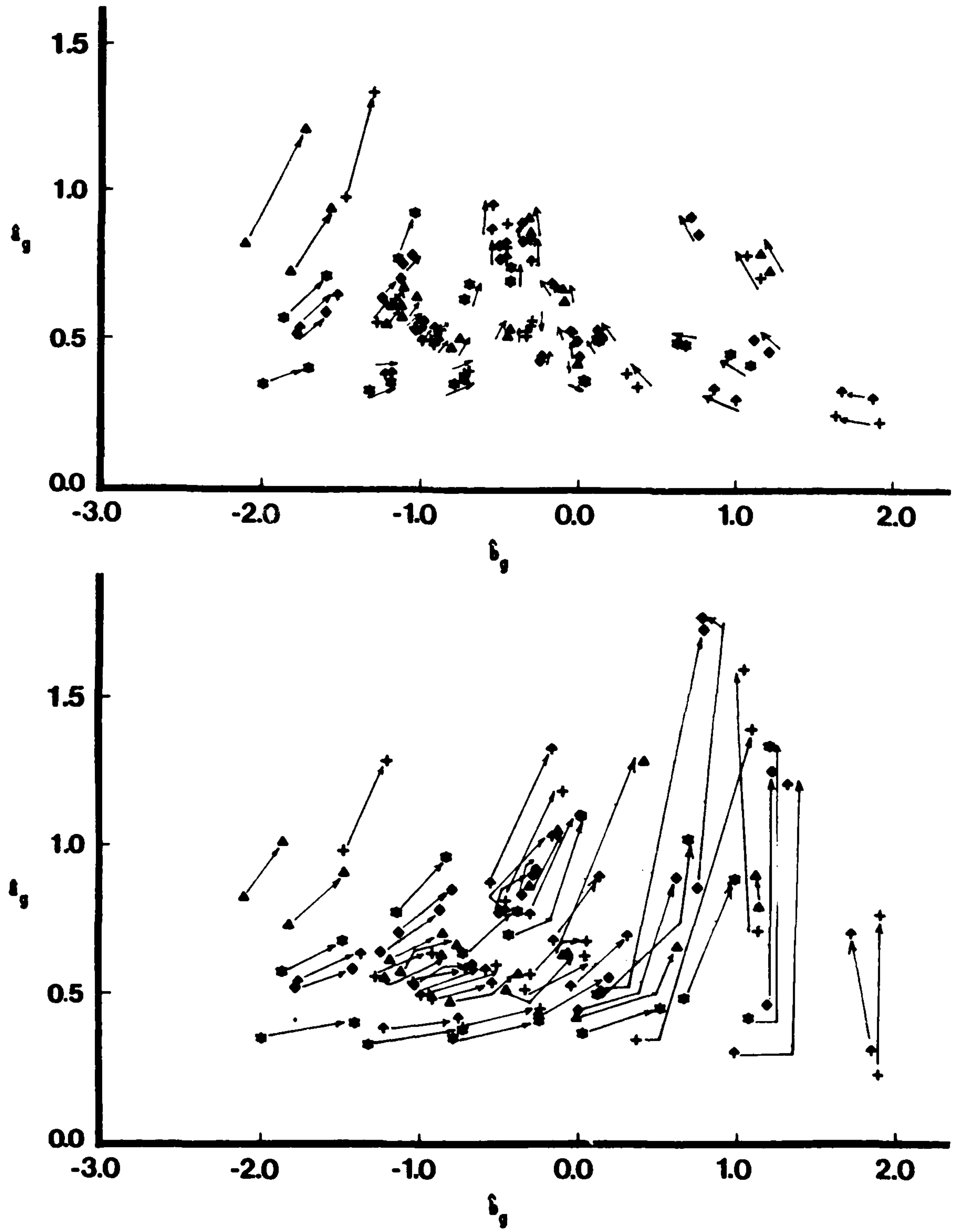


FIGURE 8-4

Estimated Item Discrimination Parameter $\hat{a}_g$ Plotted against Estimated Item DifficultyParameter $\hat{b}_g$, Which Were Obtained by the Tetrachoric Method and Those Which Were Obtained by Logist 5 Assuming Logistic Model (Upper Graph) and Three-parameter Logistic Model (Lower Graph), Respectively, for Each of the 55 Items of Test J1. For Each Item, an Arrow Is Drawn from the Tetrachoric Method Result to the Logist 5 Result. J1/1075 Case.

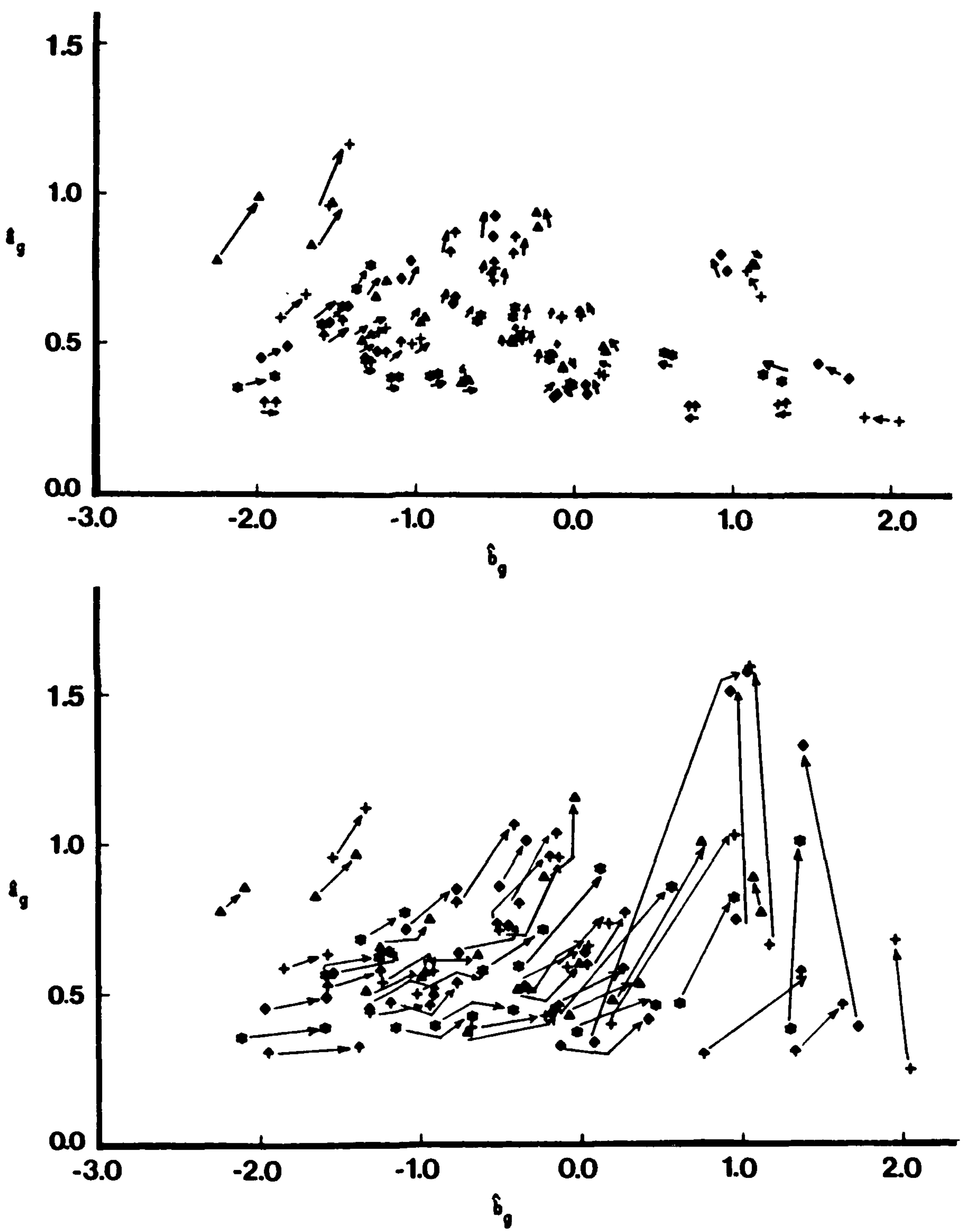


FIGURE 8-4 (Continued)

J1/2259 Case.

AD-A194 302 ADVANCEMENT OF LATENT TRAIT THEORY(U) TENNESSEE UNIV 2/2
KNOXVILLE DEPT OF PSYCHOLOGY F SAMEJIMA FEB 88
N00014-81-C-0569

UNCLASSIFIED F/G 5/8 NL

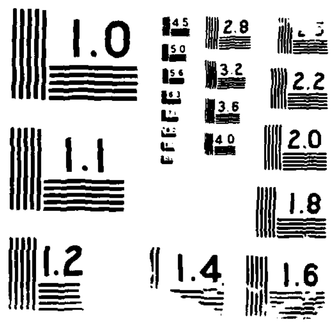

1.0
4.5
5.0
5.6
2.8
3.2
3.6
4.0
2.2
2.0
1.1
1.8
1.2
1.4
1.6

TABLE 8-2

Estimated Mean and Standard Deviation of the Distribution of $\hat{\theta}$ for Each Examinee Group Shown as Examine Group 2, for Which Logist 5 Was Used. Examinee Group 1 Indicates the Group upon Which Tetrachoric Method Was Applied and Whose Result Was Paired with the Result of Examinee Group 2.

Examinee Group 1	Examinee Group 2	Mean	S.D.
A5/0599	A5-A6	-0.67690629289170 00	0.94984351209870 00
A6/0412	A5-A6	-0.75671805937970 00	0.94412857276250 00
J1/0614	A5-A6	-0.75407910857040 00*	0.89716114961580 00*
Weighted Average		-0.70943096924490 00	0.94751457539590 00
J1/0614	J1-J2	0.77903022642650 00	0.13767142379140 01
J2/0758	J1-J2	0.81977477492590 00	0.14062298083840 01
Weighted Average		0.80154069855660 00	0.13930209452140 01
A5/0599	A5-A6-J1-J2	0.18832703348690 00	0.12160206299810 01
A6/0412	A5-A6-J1-J2	0.15141227900610 00	0.12851182990060 01
J1/0614	A5-A6-J1-J2	0.17042626954680 00	0.15224560569890 01
J2/0758	A5-A6-J1-J2	0.19383911617830 00	0.16011460339090 01
Weighted Average		0.17908582944780 00	0.14294258536480 01
J1/1075	J1/1075	0.30175006749560 00	0.12059537818110 01
J1/0614	J1/1075	0.29702633552660 00	0.12020065561900 01
Weighted Average		0.30003285528190 00	0.12045188519520 01
J1/2259	J1/2259	0.30186763094760 00	0.10778066943580 01
J1/0614	J1/2259	0.30316841351640 00	0.10807682135930 01
Weighted Average		0.30214562624770 00	0.10784396121480 01

* This is based on only 14 items while all the other results are based on at least 36 items. This result was excluded from the computation of the weighted average.

These two findings seem to justify the second scale adjustment, and also to support the consistency in the results of the two methods, i.e., Tetrachoric Method and Logist 5.

Figures 8-9 and 8-10 present the results of J1/1075 and J1/2259 Cases for each of four items of Test J1. Again in each of these graphs a solid curve represents the estimated item characteristic function in the normal ogive model, whose item parameters were originally obtained by the Tetrachoric Method and then adjusted. The other four curves are based upon the estimated item parameters obtained by Logist 5 , with two of them by assuming (two-parameter) logistic model and the other two by assuming three-parameter logistic model.

From these results we can find the following.

(3) For many items the logistic curve obtained with the second scale adjustment, which is shown by a short dashed line, is very close to the normal ogive curve, which is drawn by a solid line.

(4) For certain items, the three-parameter logistic curves are drastically different from the other three curves, whereas for many other items they are close to the other three for the range of θ , $(-1.0, \infty)$, regardless of the fact that the estimated discrimination parameters are substantially larger.

[VIII.3] Discussion

Tetrachoric Method has been criticized recently for such reasons that: 1) the tetrachoric correlation matrix does not turn out to be positive definite, and 2) the correlation does not handle too difficult and too easy items well. While the second criticism makes sense, the principal investigator does not agree with this negative standpoint. First of all, it should be remembered that, in using the tetrachoric correlation coefficient, we need the assumption of bivariate normality for each pair of response tendencies. Care must be taken, therefore, to make sure that our subjects are a practically randomly selected sample of a "non-restricted" population, before we use the Tetrachoric Method. Secondly, the first criticism is mostly based upon results obtained by poorly written computer programs of tetrachoric correlation coefficient. With a well written computer program and a suitable group of subjects, the method can be useful, unless the test includes so many too difficult and/or too easy items. This was proved by the success of the method in analyzing Shiba's Data, and also in analyzing the Iowa Test Data, which were introduced in Chapter VII.

References

[1] Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick; *Statistical theories of mental test scores.* Addison-Wesley, 1968, Chapters 17-20.

[2] Deming, W. E. *Statistical adjustment of Data.* New York: John Wiley & Sons, Inc., 1943.

[3] Lord, F. M. An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 1968, 28, 989-1020.

[4] Lord. F. M. *Applications of item response theory to practical testing problems.* Hillsdale, N. J.: Lawrence Erlbaum, 1980.

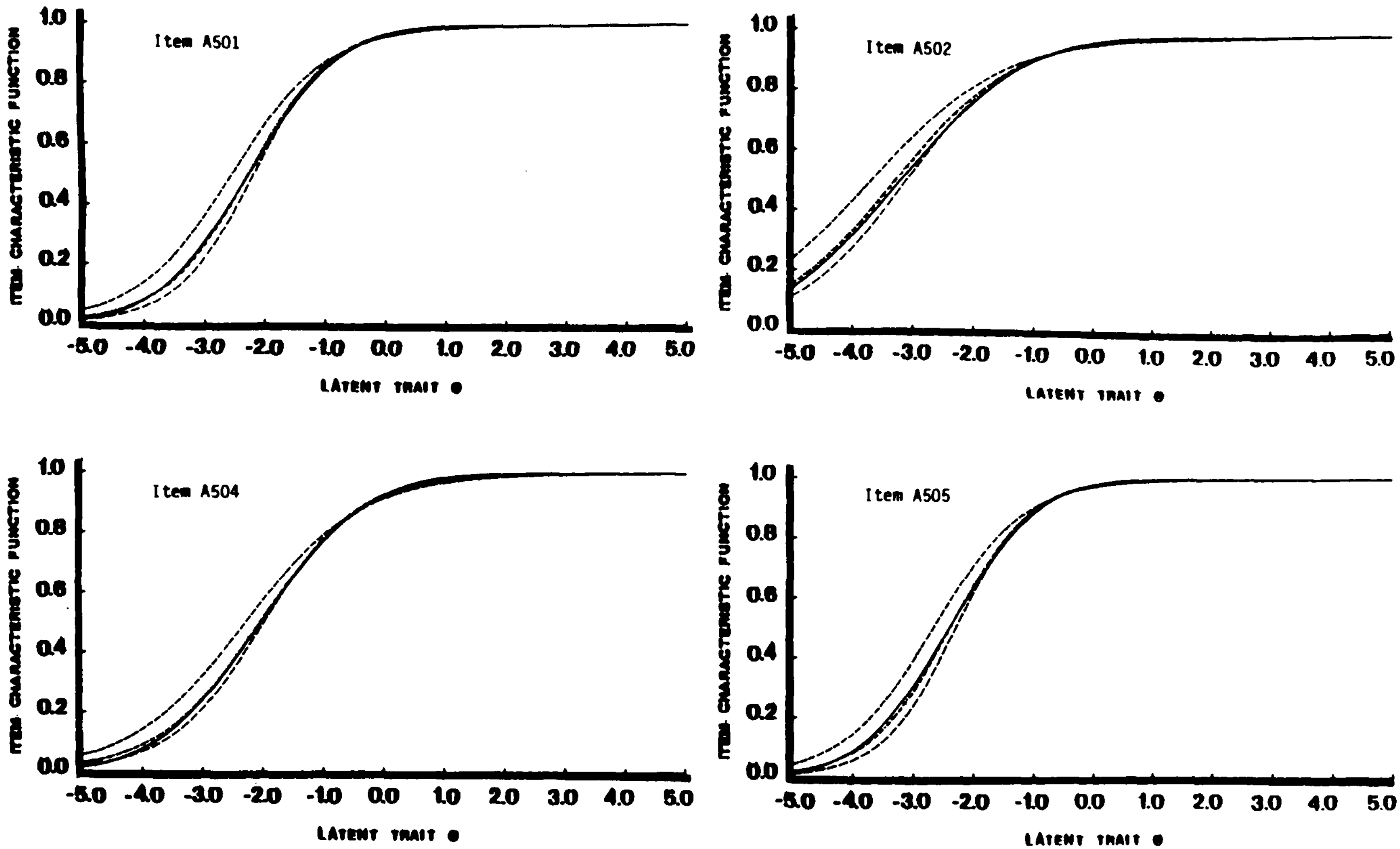


FIGURE 8-5

Estimated Item Characteristic Function in the Normal Ogive Model Obtained by the Tetrachoric Method (Solid Line), Four Estimated Item Characteristic Functions Following the Logistic Model Obtained by Using Logist 5, Two of Which Are the Results of the A5-A6 Case Based upon the First Scale Adjustment (Long Dashed Line) and upon the Second Scale Adjustment (Short Dashed Line), and the Other Two of Which Are the Results of the A5-A6-J1-J2 Case Based upon the First Scale Adjustment (Dashed Line of Medium Length) and upon the Second Scale Adjustment (Dotted Line), for Each of Four Items of Test A5.

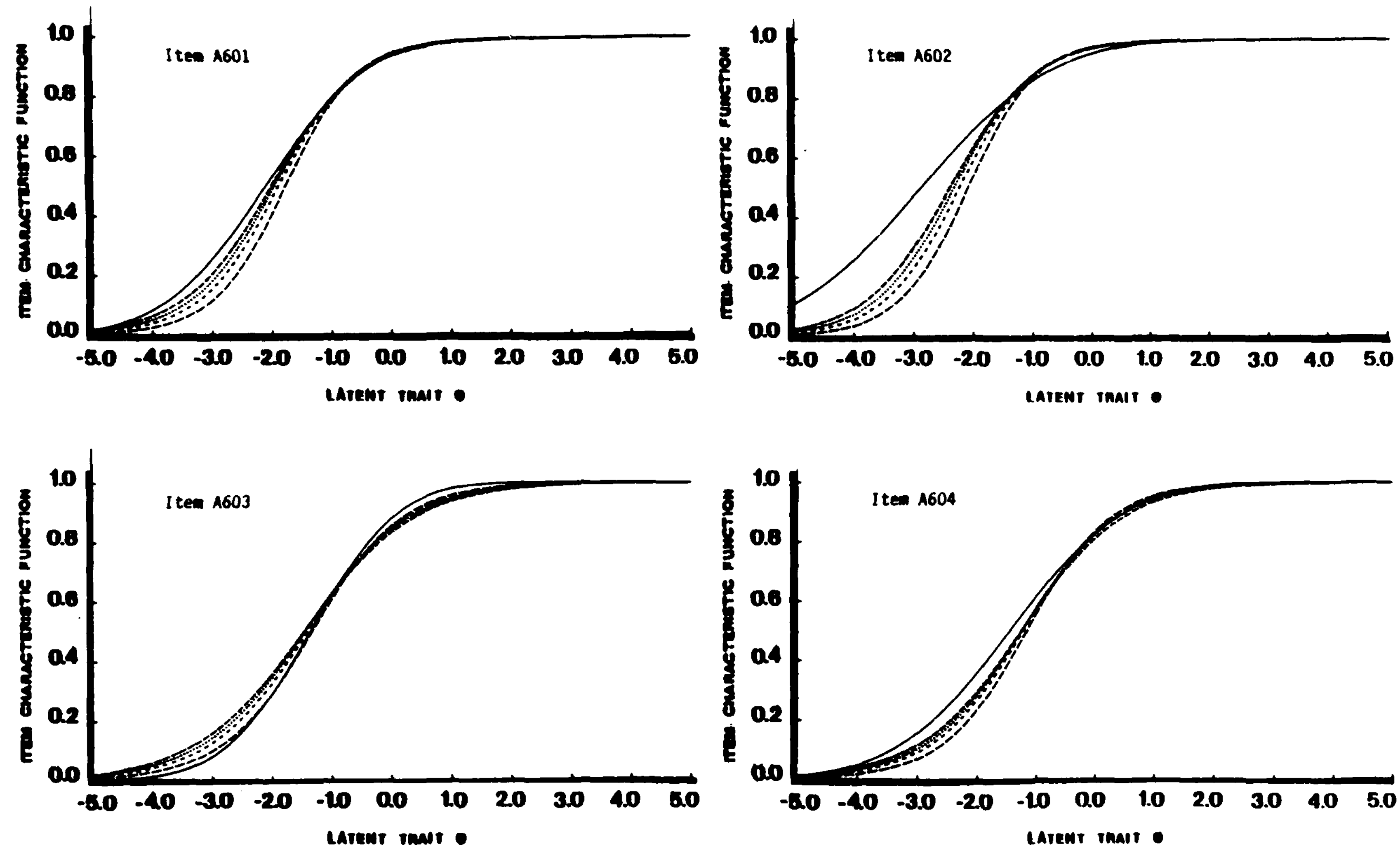


FIGURE 8-6

Estimated Item Characteristic Function in the Normal Ogive Model Obtained by the Tetrachoric Method (Solid Line), Four Estimated Item Characteristic Functions Following the Logistic Model Obtained by Using Logist 5, Two of Which Are the Results of the A5-A6 Case Based upon the First Scale Adjustment (Long Dashed Line) and upon the Second Scale Adjustment (Short Dashed Line), and the Other Two of Which Are the Results of the A5-A6-J1-J2 Case Based upon the First Scale Adjustment (Dashed Line of Medium Length) and upon the Second Scale Adjustment (Dotted Line), for Each of Four Items of Test A6.

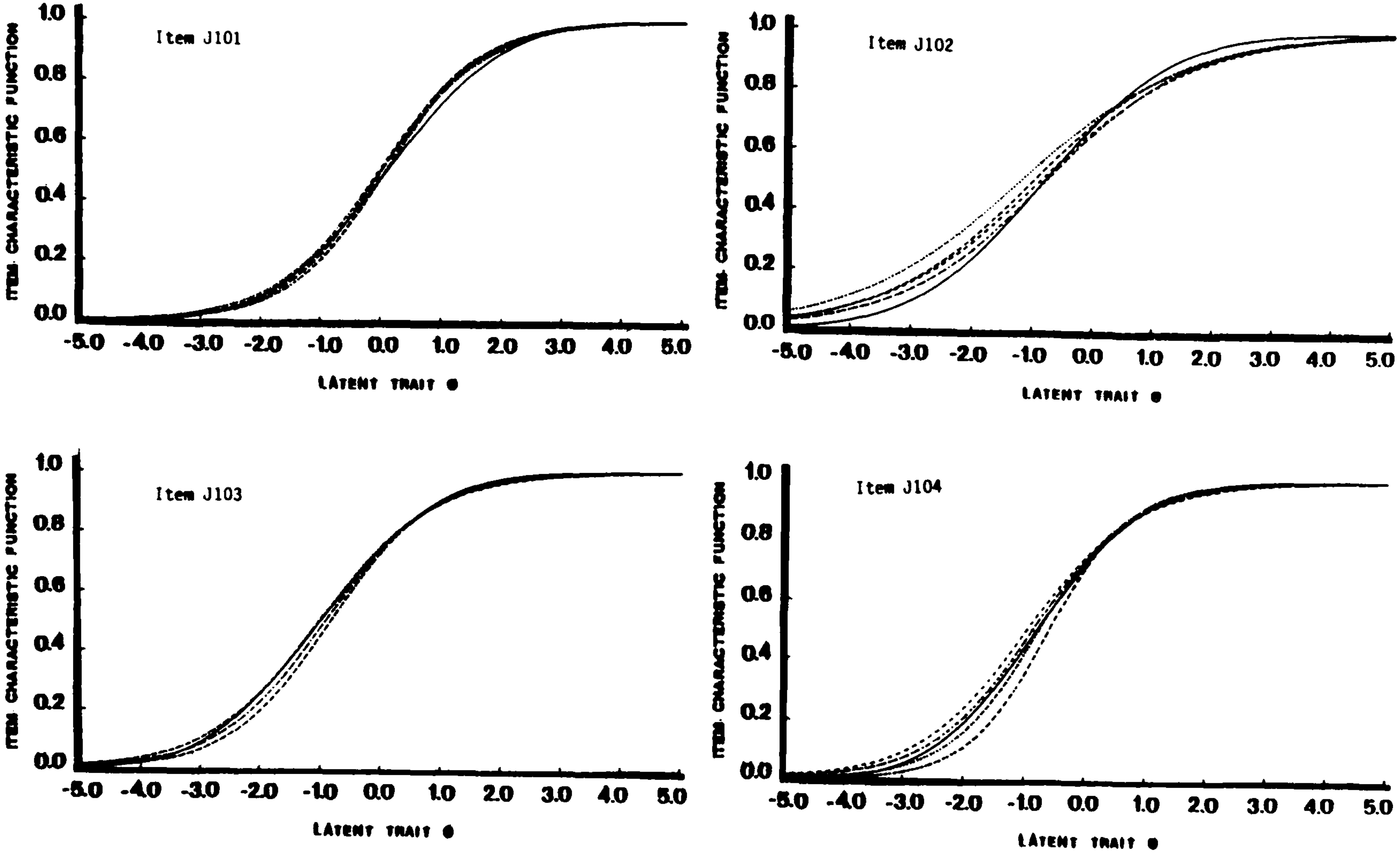


FIGURE 8-7

Estimated Item Characteristic Function in the Normal Ogive Model Obtained by the Tetrachoric Method (Solid Line), Four Estimated Item Characteristic Functions Following the Logistic Model Obtained by Using Logist 5, Two of Which Are the Results of the J1-J2 Case Based upon the First Scale Adjustment (Long Dashed Line) and upon the Second Scale Adjustment (Short Dashed Line), and the Other Two of Which Are the Results of the A5-A6-J1-J2 Case Based upon the First Scale Adjustment (Dashed Line of Medium Length) and upon the Second Scale Adjustment (Dotted Line), for Each of Four Items of Test J1.

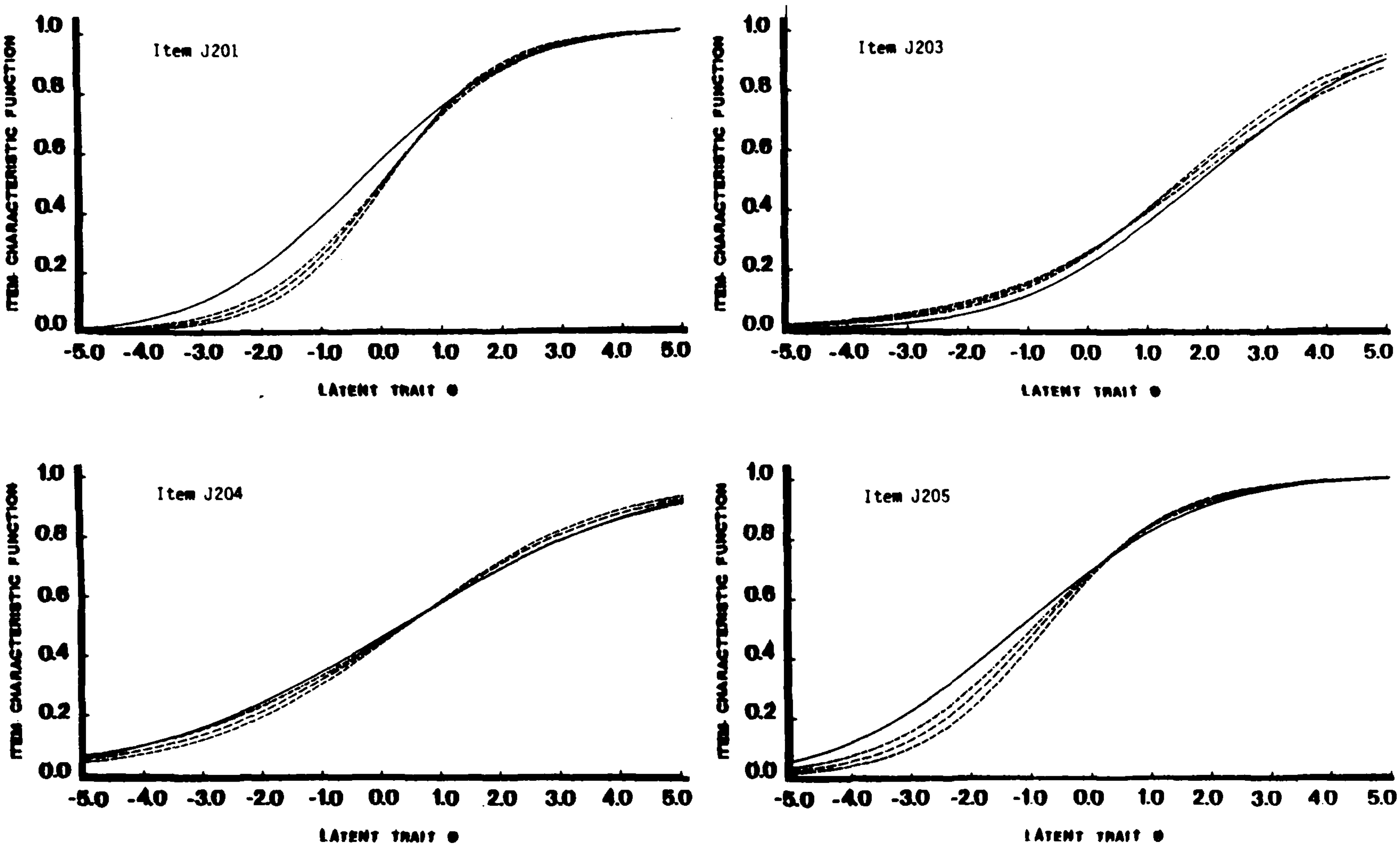


FIGURE 8-8

Estimated Item Characteristic Function in the Normal Ogive Model Obtained by the Tetrachoric Method (Solid Line), Four Estimated Item Characteristic Functions Following the Logistic Model Obtained by Using Logist 5, Two of Which Are the Results of the J1-J2 Case Based upon the First Scale Adjustment (Long Dashed Line) and upon the Second Scale Adjustment (Short Dashed Line), and the Other Two of Which Are the Results of the A5-A6-J1-J2 Case Based upon the First Scale Adjustment (Dashed Line of Medium Length) and upon the Second Scale Adjustment (Dotted Line), for Each of Four Items of Test J2.

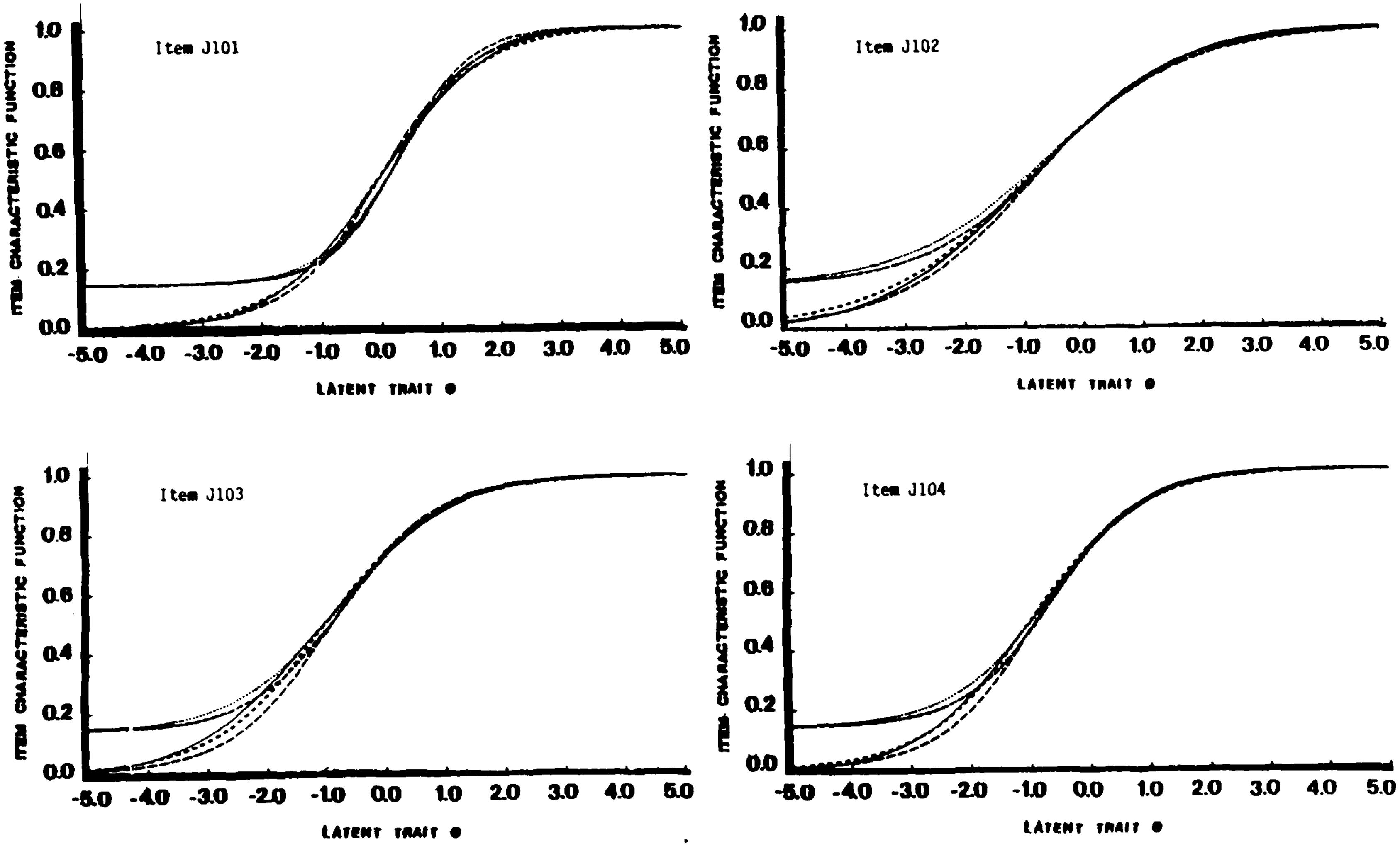


FIGURE 8-9

Estimated Item Characteristic Function in the Normal Ogive Model Obtained by the Tetrachoric Method (Solid Line), Two Estimated Item Characteristic Functions Following the Logistic Model Obtained by Using Logist 5, Which Are Based upon the First Scale Adjustment (Long Dashed Line) and upon the Second Scale Adjustment (Short Dashed Line), And Two Estimated Item Characteristic Functions Following the Three-parameter Logistic Model Obtained by Using Logist 5, Which Are Based upon the First Scale Adjustment (Dashed Line of Medium Length) and upon the Second Scale Adjustment (Dotted Line), for Each of Four Items of Test J1. Results of the J1/1075 Case.

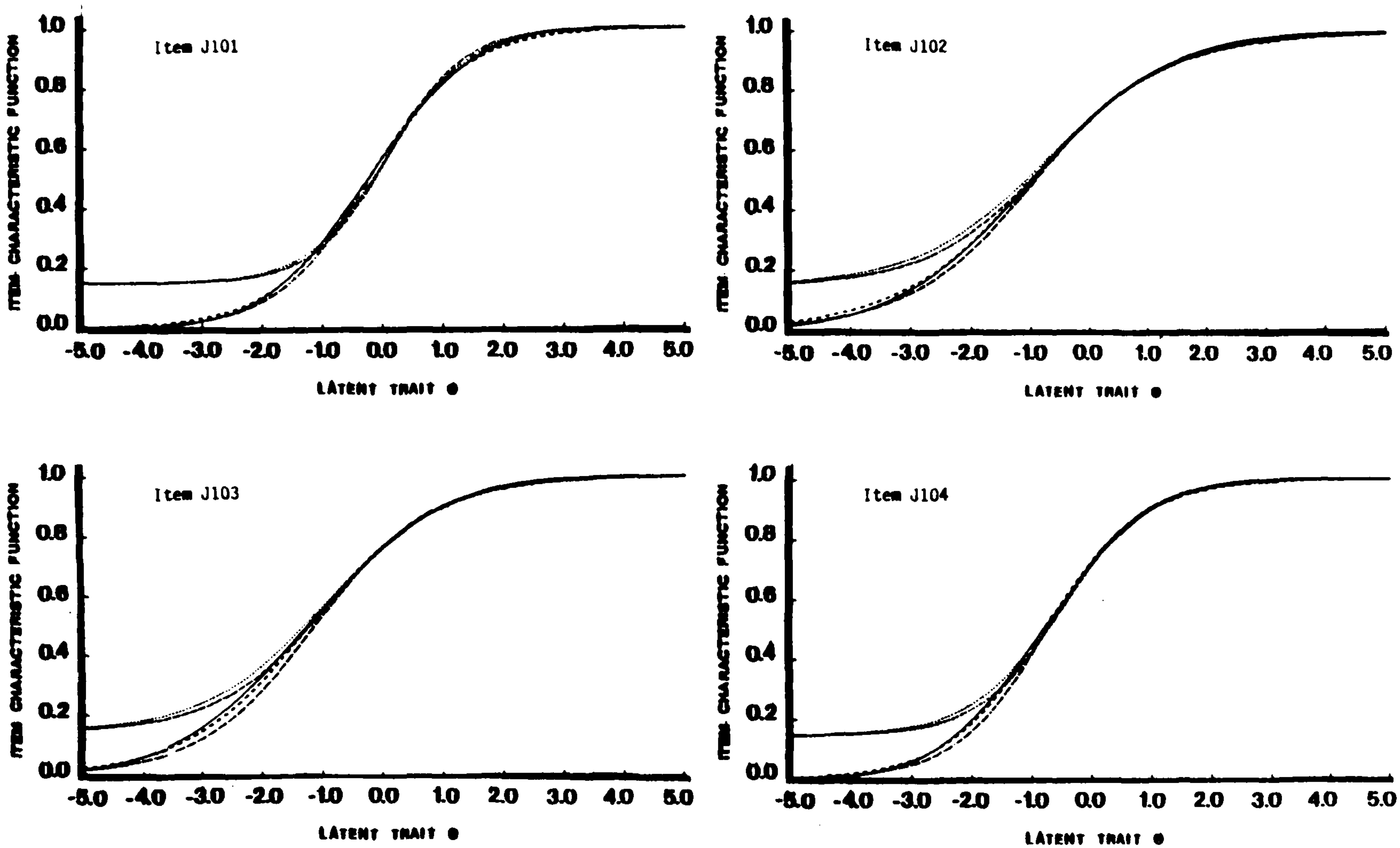


FIGURE 8-10

Estimated Item Characteristic Function in the Normal Ogive Model Obtained by the Tetrachoric Method (Solid Line), Two Estimated Item Characteristic Functions Following the Logistic Model Obtained by Using Logist 5, Which Are Based upon the First Scale Adjustment (Long Dashed Line) and upon the Second Scale Adjustment (Short Dashed Line), and Two Estimated Item Characteristic Functions Following the Three-Parameter Logistic Model Obtained by Using Logist 5, Which are Based upon the First Scale Adjustment (Dashed Line of Medium Length) and upon the Second Scale Adjustment (Dotted Line), for Each of Four Items of Test J1. Results of the J1/2259 Case.

[5] Rasch, G. *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danmarks Paedogogishe Institut, 1960.

[6] Samejima, F. Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17, 1969.

[7] Samejima, F. A general model for free-response data. *Psychometrika Monograph*, No. 18, 1972.

[8] Samejima, F. A comment on Birnbaum's three-parameter logistic model in the latent trait theory. *Psychometrika*, 1973, 38, 2, 221-233.

[9] Samejima, F. Research on the multiple-choice test item in Japan: Toward the validation of mathematical models. *ONR-Tokyo, Scientific Monograph* 3, April 1980.

[10] Samejima, F. Plausibility functions of Iowa vocabulary test items estimated by the Simple Sum Procedure of the Conditional P.D.F. Approach. *Office of Naval Research Report, 84-1*, 1984.

[11] Samejima, F. Results of item parameter estimation using Logist 5 on simulated data. *Office of Naval Research Report, 84-3*, 1984.

[12] Shiba, S. Construction of a scale for acquisition of word meanings. *Bulletin of Faculty of Education, University of Tokyo*, 17, 1978. (in Japanese)

[13] Wingersky, M. S., M. A. Barton, and F. M. Lord. *Logist User's Guide.* Princeton: Educational Testing Service, 1982.

IX Item Parameter Estimation Using Logist 5 on Simulated Data

It has been shown in the result of the item analysis of Shiba's Data that the discrimination parameter estimated by Logist 5 tends to be greater, if the third parameter c_g is set free, and to a lesser extent the same is true with the difficulty parameter. It has also been observed that in spite of this fact the resulting estimated item characteristic function tends to be closer to those obtained by the Tetrachoric Method and by Logist 5 with zero as the set value of c_g, than those enhanced $\hat{a}_g$ and $\hat{b}_g$ suggest. This fact should be taken as a warning to researchers who have been accepting $\hat{a}_g$ as the discrimination power and $\hat{b}_g$ as the difficulty index of the item when three-parameter logistic model is assumed. The truth is that, unlike in the two-parameter model such as the normal ogive and the logistic models, in the three-parameter model both the discrimination and the difficulty of the item are contaminated by the guessing parameter c_g, and neither a_g nor b_g has a meaning by itself.

Since Shiba's Data are empirical data, there is no way of knowing the true item characteristic function of each item. With simulated data, however, we can produce items whose *true* item characteristic functions follow the normal ogive model. If we assume the three-parameter logistic model for these test items, instead of the normal ogive or the logistic model and estimate the three item parameters simultaneously by using Logist 5, shall we obtain the estimate of c_g which is close enough to the true value, zero, and the estimates of a_g and b_g which are close enough to their true values? Or will this additional free parameter c_g contaminate the result so that we will be provided with a substantially different set of estimated item parameters?

In this part of the research we tried Logist 5 on a set of simulated data and pursued this issue. This chapter will outline its method and results. For the details and more information, see [I.2.8].

[IX.1] Simulated Data

Two tests were hypothesized, which consist of ten and thirty-five dichotonmous items, respectively, each following the normal ogive model. For brevity, we shall call them Ten Item Test and Thirty-Five Item Test, respectively. They were used separately, and also together as a test of forty-five items, and these cases are called Cases 1, 2 and 3, respectively. In addition to these three, we have Case 4 of eighty items. This last case was created rather artificially in order to observe the results based upon a larger number of items.

The hypothesized ability distribution is uniform, for the interval of ability θ, (-2.5, 2.5). Starting with -2.475 and ending with -2.475, five hypothetical subjects were placed at each of the one hundred points with the common interval width of 0.050, to create the 500 Subject Case. Later, this was repeated three times more, to obtain the 2,000 Subject Case. Monte Carlo Method was used to produce a response pattern for each hypothetical subject. In practice, however, each item of the Thirty-Five Item Test was a graded item having two difficulty parameters. It was redichotomized by using the first difficulty parameter only, and each response pattern was adjusted accordingly. Later, when Case 4 was created, these same response patterns were used again by redichotomozing each item using the second difficulty parameter.

[IX.2] Method

Logist 5 was used twice for each combination of a subject group and a set of items, first for estimating the three parameters a_g , b_g and c_g , and then for estimating a_g and b_g only by setting $c_g = 0$, as we did in analyzing Shiba's Data. This means that we assumed the three-parameter logistic model in the first situation, and the (two-parameter) logistic model in the second.

In Logist 5, the origin and the unit of ability θ are set at the mean and the standard deviation of its maximum likelihood estimate $\hat{\theta}_s$ for all subjects whose $\hat{\theta}_s$ are within the interval of (-3.0, 3.0) as the result of the last iteration. This causes some problems, because in so doing the effect of the error involved in $\hat{\theta}_s$ is ignored, and also the excluded subjects affect the resulting scale in each case. Since there is no simple way to make the scale adjustment, however, these scale differences were not adjusted in the research.

The theoretical item and individual parameters were transformed in order to make them comparable to the results obtained by Logist 5. Since we have

$$E(\theta) = (\alpha + \beta)/2 \tag{9.1}$$

and

$$Var.(\theta) = (\beta - \alpha)^2/12 \tag{9.2}$$

for the uniform distribution with the interval (α, β) , the origin of θ was kept as it was, and the unit was made 1.443375673 times larger.

[IX.3] Results

Tables 9-1 through 9-8 present the true and estimated discrimination and difficulty parameters for the eight combinations of a hypothetical test and a subject group. We can see there is a general tendency of the enhancement of the estimated discrimination parameters, and to a lesser extent of the estimated difficulty parameters.

This enhancement of the estimated discrimination parameters is more revealing if we plot those values against the theoretical discrimination parameters a_g . Figures 9-1 and 9-2 illustrate those results of the 500 and 2,000 Subject Cases, respectively, with both the three-parameter logistic and the logistic models assumed for Case 3. In each of these graphs, the upper limit set for the estimated discrimination parameter in using Logist 5 is indicated by a dotted, horizontal line. Also a solid line with the angle of 45 degrees from the abscissa passing (0,0) is drawn in each graph. We can see in these results that, although there is some improvement in those obtained by setting $c_g = 0.0$ in Logist 5, many discrimination parameters are outrageously overestimated in both the 500 and 2,000 Cases. This tendency is even more conspicuous for the two cases of smaller number of test items. The enhancement is reduced to some extent in the results of Case 4, especially in the 2,000 Subject Case. It is still substantial, however.

The corresponding results on the difficulty parameters in the 500 and 2,000 Subject Cases are illustrated as Figures 9-3 and 9-4, respectively. In these graphs, the interval of θ, $(-\sqrt{3}, \sqrt{3})$, for

TABLE 9-1

Theoretical and Estimated Item Parameters in the 500 and 2,000 Ten Item Test. Three-Parameter Logistic Model Is Assumed. Case 1.

Item	Discrimination Parameter				Difficulty Parameter				Guessing Parameter	
	Theoretical		Estimated		Theoretical		Estimated		Estimated	
	Org.	Adj.	500 S.C.	2,000 S.C.	Org.	Adj.	500 S.C.	2,000 S.C.	500 S.C.	2,000 S.C.
1	1.50000	2.16506	4.00000	0.95615	-2.50000	-1.73205	-7.00850	-4.53097	0.11111	0.00000
2	1.00000	1.44338	1.12847	1.89857	-2.00000	-1.38564	-2.27705	-1.25111	0.11111	0.23137
3	2.50000	3.60844	4.00000	7.00000	-1.50000	-1.03923	-1.49140	-1.02797	0.01285	0.01897
4	1.00000	1.44338	1.49368	2.16571	-1.00000	-0.69282	-0.76647	-0.34862	0.12154	0.15474
5	1.50000	2.16506	4.00000	5.73384	-0.50000	-0.34641	-0.39155	-0.07764	0.01397	0.04219
6	1.00000	1.44338	1.81350	2.45766	0.00000	0.00000	-0.00846	0.25474	0.04756	0.06358
7	2.00000	2.88675	3.08340	6.30172	0.50000	0.34641	0.32296	0.49745	0.00000	0.00499
8	1.00000	1.44338	1.47734	2.11129	1.00000	0.69282	0.72304	0.81349	0.00000	0.00645
9	2.00000	2.88675	4.00000	5.73230	1.50000	1.03923	1.10169	1.04610	0.00000	0.00000
10	1.00000	1.44338	0.65759	0.98121	2.00000	1.38564	2.60342	2.08713	0.00000	0.00000

TABLE 9-2

Theoretical and Estimated Item Parameters in the 500 and 2,000 Subject Cases for Each Item of the Ten Item Test. Logistic Model Is Assumed. Case 1.

Item	Discrimination Parameter				Difficulty Parameter			
	Theoretical		Estimated		Theoretical		Estimated	
	Org.	Adj.	500 S.C.	2,000 S.C.	Org.	Adj.	500 S.C.	2,000 S.C.
1	1.50000	2.16506	7.00000	4.46012	-2.50000	-1.73205	-1.79784	-1.82777
2	1.00000	1.44338	1.66645	1.48512	-2.00000	-1.38564	-1.40424	-1.41994
3	2.50000	3.60844	6.83938	4.77435	-1.50000	-1.03923	-1.00147	-0.98488
4	1.00000	1.44338	1.38928	1.55463	-1.00000	-0.69282	-0.64011	-0.63939
5	1.50000	2.16506	3.61872	3.26944	-0.50000	-0.34641	-0.18574	-0.25476
6	1.00000	1.44338	1.53595	1.60692	0.00000	0.00000	0.11857	0.10502
7	2.00000	2.88675	3.38480	4.45453	0.50000	0.34641	0.52054	0.51174
8	1.00000	1.44338	1.70183	1.66029	1.00000	0.69282	0.86021	0.88896
9	2.00000	2.88675	7.00000	6.08234	1.50000	1.03923	1.18334	1.18419
10	1.00000	1.44338	1.00086	1.12968	2.00000	1.38564	2.09119	2.01825

TABLE 9-3

Theoretical and Estimated Item Parameters in the 500 and 2,000 Subject Cases for Each Item of the Thirty-Five Item Test. Three-parameter Logistic Model Is Assumed. Case 2.

Item	Discrimination Parameter				Difficulty Parameter				Guessing Parameter	
	Theoretical		Estimated		Theoretical		Estimated		Estimated	
	Org.	Adj.	500 S.C.	2,000 S.C.	Org.	Adj.	500 S.C.	2,000 S.C.	500 S.C.	2,000 S.C.
11	1.80000	2.59808	---	---	-4.75000	-3.29090	---	---	---	---
12	1.90000	2.74241	---	---	-4.50000	-3.11769	---	---	---	---
13	2.00000	2.88675	---	---	-4.25000	-2.94449	---	---	---	---
14	1.50000	2.16506	7.00000	2.10115	-4.00000	-2.77128	-2.31556	-3.02161	0.08667	0.02122
15	1.60000	2.30940	---	4.95191	-3.75000	-2.59808	---	-2.41505	---	0.02122
16	1.40000	2.02073	7.00000	2.01814	-3.50000	-2.42487	-2.31184	-2.66637	0.08667	0.02122
17	1.90000	2.74241	2.34600	6.25083	-3.00000	-2.07846	-2.49184	-1.97065	0.08667	0.00000
18	1.80000	2.59808	1.66937	2.45684	-3.00000	-2.07846	-2.53506	-2.23919	0.08667	0.02122
19	1.60000	2.30940	5.95996	2.62153	-2.75000	-1.90526	-1.63507	-1.90260	0.33907	0.02122
20	2.00000	2.88675	4.54064	3.93175	-2.50000	-1.73205	-1.75362	-1.66113	0.00000	0.16847
21	1.50000	2.16506	3.60897	3.57269	-2.25000	-1.55885	-1.52887	-1.47593	0.00000	0.00000
22	1.70000	2.45374	3.72440	5.79677	-2.00000	-1.38564	-1.26918	-1.20411	0.19618	0.17118
23	1.50000	2.16506	5.93530	4.14441	-1.75000	-1.21244	-0.86540	-1.04556	0.31164	0.12398
24	1.40000	2.02073	2.15613	2.50797	-1.50000	-1.03923	-0.91905	-0.92924	0.05498	0.08893
25	2.00000	2.88675	4.30412	3.82653	-1.25000	-0.86603	-0.82709	-0.84287	0.00000	0.01227
26	1.60000	2.30940	3.08081	3.00092	-1.00000	-0.69282	-0.61492	-0.62173	0.00000	0.03369
27	1.80000	2.59808	3.41671	3.29292	-0.75000	-0.51962	-0.42482	-0.50742	0.01715	0.00712
28	1.70000	2.45374	3.73975	3.15448	-0.50000	-0.34641	-0.29610	-0.32412	0.00000	0.00000
29	1.90000	2.74241	3.55043	3.43681	-0.25000	-0.17321	-0.16918	-0.15937	0.00000	0.00000
30	1.70000	2.45374	3.26689	2.97724	0.00000	0.00000	0.06045	0.03331	0.00000	0.00000
31	1.50000	2.16506	2.51771	2.68637	0.25000	0.17321	0.24396	0.17012	0.00000	0.00132
32	1.80000	2.59808	3.36240	3.58997	0.50000	0.34641	0.38959	0.35692	0.00000	0.00000
33	1.40000	2.02073	2.49348	2.33960	0.75000	0.51962	0.49717	0.47978	0.00000	0.00000
34	1.90000	2.74241	3.92561	3.63878	1.00000	0.69282	0.64148	0.65455	0.00000	0.00000
35	2.00000	2.88675	4.36628	3.48851	1.25000	0.86603	0.78414	0.85172	0.00000	0.00000
36	1.60000	2.30940	2.61472	2.72612	1.50000	1.03923	1.01803	1.02596	0.00000	0.00000
37	1.70000	2.45374	2.54991	2.90554	1.75000	1.21244	1.21299	1.22349	0.00000	0.00000
38	1.40000	2.02073	1.97089	2.25224	2.00000	1.38564	1.38765	1.34179	0.00000	0.00000
39	1.90000	2.74241	6.12277	3.14405	2.25000	1.55885	1.47525	1.52861	0.07920	0.00000
40	1.60000	2.30940	4.30567	3.39256	2.50000	1.73205	1.51061	1.62665	0.00000	0.00000
41	1.50000	2.16506	2.50260	2.18581	2.75000	1.90526	2.09183	1.97325	0.00000	0.00000
42	1.70000	2.45374	7.00000	6.00998	3.00000	2.07846	1.82767	1.88083	0.00000	0.00000
43	1.80000	2.59808	3.18005	1.91086	3.25000	2.25167	2.10989	2.68272	0.00000	0.00000
44	2.00000	2.88675	---	---	3.50000	2.42487	---	---	---	---
45	1.40000	2.02073	---	7.00000	3.75000	2.59808	---	2.19254	---	0.00198

TABLE 9-4

Theoretical and Estimated Item Parameters in the 500 and 2,000 Subject Cases for Each Item of the Thirty-Five Item Test. Logistic Model Is Assumed. Case 2.

Item	Discrimination Parameter				Difficulty Parameter			
	Theoretical		Estimated		Theoretical		Estimated	
	Org.	Adj.	500 S.C.	2,000 S.C.	Org.	Adj.	500 S.C.	2,000 S.C.
11	1.80000	2.59808	---	---	-4.75000	-3.29090	---	---
12	1.90000	2.74241	---	---	-4.50000	-3.11769	---	---
13	2.00000	2.88675	---	---	-4.25000	-2.94449	---	---
14	1.50000	2.16506	7.00000	2.66184	-4.00000	-2.77128	-2.24211	-2.72393
15	1.60000	2.30940	---	1.80462	-3.75000	-2.59808	---	-3.48102
16	1.40000	2.02073	5.85000	2.18952	-3.50000	-2.42487	-2.27498	-2.57504
17	1.90000	2.74241	2.65490	4.80292	-3.00000	-2.07846	-2.38002	-1.98730
18	1.80000	2.59808	1.86159	2.27740	-3.00000	-2.07846	-2.43453	-2.28213
19	1.60000	2.30940	3.04990	2.35694	-2.75000	-1.90526	-1.84166	-1.94374
20	2.00000	2.88675	5.26700	3.67092	-2.50000	-1.73205	-1.69718	-1.71751
21	1.50000	2.16506	3.68387	3.53050	-2.25000	-1.55885	-1.51712	-1.48068
22	1.70000	2.45374	2.78792	3.44789	-2.00000	-1.38564	-1.41474	-1.32688
23	1.50000	2.16506	2.29748	3.17041	-1.75000	-1.21244	-1.21515	-1.15032
24	1.40000	2.02073	2.06001	2.20858	-1.50000	-1.03923	-0.98995	-1.02123
25	2.00000	2.88675	4.28130	3.59910	-1.25000	-0.86603	-0.85496	-0.86681
26	1.60000	2.30940	3.00410	2.66034	-1.00000	-0.69282	-0.63407	-0.66631
27	1.80000	2.59808	2.96481	3.16909	-0.75000	-0.51962	-0.45806	-0.52025
28	1.70000	2.45374	3.67673	3.11533	-0.50000	-0.34641	-0.30288	-0.32330
29	1.90000	2.74241	3.46570	3.44479	-0.25000	-0.17321	-0.17130	-0.15421
30	1.70000	2.45374	3.24069	2.95857	0.00000	0.00000	0.06386	0.04162
31	1.50000	2.16506	2.48202	2.63270	0.25000	0.17321	0.24964	0.17709
32	1.80000	2.59808	3.28184	3.56525	0.50000	0.34641	0.39818	0.36913
33	1.40000	2.02073	2.44574	2.33705	0.75000	0.51962	0.50649	0.49057
34	1.90000	2.74241	3.75665	3.62024	1.00000	0.69282	0.65520	0.66711
35	2.00000	2.88675	4.35412	3.52539	1.25000	0.86603	0.79956	0.86196
36	1.60000	2.30940	2.74792	2.81144	1.50000	1.03923	1.02172	1.03020
37	1.70000	2.45374	2.63339	3.03734	1.75000	1.21244	1.21153	1.22028
38	1.40000	2.02073	2.03766	2.36148	2.00000	1.38564	1.38059	1.33170
39	1.90000	2.74241	3.39088	3.39749	2.25000	1.55885	1.45505	1.50674
40	1.60000	2.30940	4.77467	3.54667	2.50000	1.73205	1.49452	1.60441
41	1.50000	2.16506	2.58117	2.27431	2.75000	1.90526	2.07618	1.93828
42	1.70000	2.45374	7.00000	5.94718	3.00000	2.07846	1.83612	1.85210
43	1.80000	2.59808	2.91442	1.95702	3.25000	2.25167	2.15710	2.63901
44	2.00000	2.88675	---	---	3.50000	2.42487	---	---
45	1.40000	2.02073	---	4.38154	3.75000	2.59808	---	2.24655

TABLE 9-5

Theoretical and Estimated Item Parameters in the 500 and 2,000 Subject Cases for Each Item of the Ten Item Test and the Thirty-Five Item Test. Three-Parameter Logistic Model Is Assumed. Case 3.

Item	Discrimination Parameter				Difficulty Parameter				Guessing Parameter	
	Theoretical		Estimated		Theoretical		Estimated		Estimated	
	Org.	Adj.	500 S.C.	2,000 S.C.	Org.	Adj.	500 S.C.	2,000 S.C.	500 S.C.	2,000 S.C.
1	1.50000	2.16506	7.00000	3.16857	-2.50000	-1.73205	-1.58337	-1.70186	0.31911	0.00000
2	1.00000	1.44338	3.87013	2.20593	-2.00000	-1.38564	-0.82818	-0.99213	0.42360	0.33765
3	2.50000	3.60844	6.61387	6.11440	-1.50000	-1.03923	-0.88235	-0.94942	0.07876	0.06904
4	1.00000	1.44338	1.66601	1.60446	-1.00000	-0.69282	-0.57784	-0.65341	0.06874	0.02886
5	1.50000	2.16506	2.77616	2.49209	-0.50000	-0.34641	-0.27650	-0.33620	0.00000	0.00000
6	1.00000	1.44338	1.55749	1.58121	0.00000	0.00000	0.02360	-0.00453	0.01465	0.00128
7	2.00000	2.88675	2.96359	3.40269	0.50000	0.34641	0.36510	0.36758	0.00000	0.00022
8	1.00000	1.44338	1.76403	1.68448	1.00000	0.69282	0.65648	0.69709	0.00000	0.00000
9	2.00000	2.88675	4.38452	3.69650	1.50000	1.03923	0.97398	0.98698	0.00000	0.00000
10	1.00000	1.44338	1.54466	1.67432	2.00000	1.38564	1.32856	1.35974	0.00000	0.00000
11	1.80000	2.59808	---	---	-4.75000	-3.29090	---	---	---	---
12	1.90000	2.74241	---	---	-4.50000	-3.11769	---	---	---	---
13	2.00000	2.88675	---	---	-4.25000	-2.94449	---	---	---	---
14	1.50000	2.16506	7.00000	2.19806	-4.00000	-2.77128	-2.37410	-2.95371	0.08826	0.10234
15	1.60000	2.30940	---	1.90897	-3.75000	-2.59808	---	-3.36769	---	0.10234
16	1.40000	2.02073	7.00000	1.84048	-3.50000	-2.42487	-2.37226	-2.75675	0.08826	0.10234
17	1.90000	2.74241	2.00167	5.60182	-3.00000	-2.07846	-2.63823	-1.99530	0.08826	0.00000
18	1.80000	2.59808	1.62591	2.21154	-3.00000	-2.07846	-2.57808	-2.29576	0.08826	0.10234
19	1.60000	2.30940	4.45727	2.36176	-2.75000	-1.90526	-1.63765	-1.92579	0.43106	0.10234
20	2.00000	2.88675	6.38769	3.87045	-2.50000	-1.73205	-1.73779	-1.68322	0.00000	0.14877
21	1.50000	2.16506	2.89572	3.06226	-2.25000	-1.55885	-1.57782	-1.50616	0.00000	0.00000
22	1.70000	2.45374	3.03416	4.41425	-2.00000	-1.38564	-1.33087	-1.18774	0.14331	0.20630
23	1.50000	2.16506	3.81430	3.36983	-1.75000	-1.21244	-0.89510	-1.05681	0.28760	0.10975
24	1.40000	2.02073	2.17074	2.43977	-1.50000	-1.03923	-0.87115	-0.92907	0.09159	0.08490
25	2.00000	2.88675	3.70917	3.62242	-1.25000	-0.86603	-0.81486	-0.83289	0.00000	0.01472
26	1.60000	2.30940	2.86861	2.96961	-1.00000	-0.69282	-0.59335	-0.60991	0.00000	0.03712
27	1.80000	2.59808	3.06301	3.14587	-0.75000	-0.51962	-0.41924	-0.50798	0.00089	0.00000
28	1.70000	2.45374	4.17905	3.18564	-0.50000	-0.34641	-0.27372	-0.31965	0.00000	0.00000
29	1.90000	2.74241	3.65756	3.30394	-0.25000	-0.17321	-0.15748	-0.15832	0.00000	0.00000
30	1.70000	2.45374	3.36662	2.92438	0.00000	0.00000	0.05765	0.03296	0.00000	0.00000
31	1.50000	2.16506	2.49791	2.60306	0.25000	0.17321	0.24144	0.16800	0.00000	0.00000
32	1.80000	2.59808	2.97532	3.17624	0.50000	0.34641	0.38466	0.35806	0.00000	0.00000
33	1.40000	2.02073	2.39944	2.24392	0.75000	0.51962	0.49468	0.48257	0.00000	0.00000
34	1.90000	2.74241	3.56410	3.65448	1.00000	0.69282	0.64140	0.65986	0.00000	0.00000
35	2.00000	2.88675	4.31324	3.29432	1.25000	0.86603	0.78225	0.85445	0.00000	0.00000
36	1.60000	2.30940	2.51338	2.56816	1.50000	1.03923	1.01267	1.02876	0.00000	0.00000
37	1.70000	2.45374	2.43304	2.70841	1.75000	1.21244	1.20847	1.22648	0.00000	0.00000
38	1.40000	2.02073	1.99664	2.25976	2.00000	1.38564	1.37635	1.33734	0.00000	0.00000
39	1.90000	2.74241	2.91426	3.11288	2.25000	1.55885	1.47260	1.52379	0.00000	0.00000
40	1.60000	2.30940	3.98169	3.20664	2.50000	1.73205	1.51130	1.63052	0.00000	0.00000
41	1.50000	2.16506	2.40751	2.03311	2.75000	1.90526	2.12160	2.00855	0.00000	0.00000
42	1.70000	2.45374	7.00000	6.06362	3.00000	2.07846	1.86710	1.88062	0.00000	0.00000
43	1.80000	2.59808	2.68843	1.73795	3.25000	2.25167	2.20918	2.79685	0.00000	0.00000
44	2.00000	2.88675	---	---	3.50000	2.42487	---	---	---	---
45	1.40000	2.02073	---	7.00000	3.75000	2.59808	---	2.19260	---	0.00198

TABLE 9-6

Theoretical and Estimated Item parameters in the 500 and 2,000 Subject Cases for Each Item of the Ten Item Test and the Thirty-Five Item Test. Logistic Model Is Assumed. Case 3.

Item	Discrimination Parameter				Difficulty Parameter			
	Theoretical		Estimated		Theoretical		Estimated	
	Org.	Adj.	500 S.C.	2,000 S.C.	Org.	Adj.	500 S.C.	2,000 S.C.
1	1.50000	2.16506	4.95337	3.28161	-2.50000	-1.73205	-1.67240	-1.67029
2	1.00000	1.44338	1.69224	1.57696	-2.00000	-1.38564	-1.36571	-1.35902
3	2.50000	3.60844	4.07728	4.17736	-1.50000	-1.03923	-1.01281	-1.02694
4	1.00000	1.44338	1.46003	1.53778	-1.00000	-0.69282	-0.70019	-0.70004
5	1.50000	2.16506	2.68987	2.47682	-0.50000	-0.34641	-0.30131	-0.34419
6	1.00000	1.44338	1.44917	1.55961	0.00000	0.00000	-0.01392	-0.00994
7	2.00000	2.88675	2.86243	3.32482	0.50000	0.34641	0.36627	0.36951
8	1.00000	1.44338	1.72479	1.67575	1.00000	0.69282	0.66329	0.70095
9	2.00000	2.88675	4.34492	3.69979	1.50000	1.03923	0.98918	0.99343
10	1.00000	1.44338	1.53441	1.68119	2.00000	1.38564	1.34230	1.36275
11	1.80000	2.59808	---	---	-4.75000	-3.29090	---	---
12	1.90000	2.74241	---	---	-4.50000	-3.11769	---	---
13	2.00000	2.88675	---	---	-4.25000	-2.94449	---	---
14	1.50000	2.16506	7.00000	2.91928	-4.00000	-2.77128	-2.15752	-2.60081
15	1.60000	2.30940	---	1.74182	-3.75000	-2.59808	---	-3.53483
16	1.40000	2.02073	7.00000	2.09438	-3.50000	-2.42487	-2.15752	-2.60017
17	1.90000	2.74241	2.73182	5.71243	-3.00000	-2.07846	-2.32255	-1.91514
18	1.80000	2.59808	1.93571	2.41736	-3.00000	-2.07846	-2.37937	-2.22083
19	1.60000	2.30940	2.78781	2.41276	-2.75000	-1.90526	-1.85334	-1.92091
20	2.00000	2.88675	7.00000	3.84414	-2.50000	-1.73205	-1.66676	-1.69924
21	1.50000	2.16506	3.05070	3.28211	-2.25000	-1.55885	-1.54605	-1.48938
22	1.70000	2.45374	2.81049	3.11682	-2.00000	-1.38564	-1.42047	-1.33659
23	1.50000	2.16506	2.18879	2.84317	-1.75000	-1.21244	-1.22508	-1.15676
24	1.40000	2.02073	1.98367	2.16013	-1.50000	-1.03923	-0.99519	-1.02270
25	2.00000	2.88675	3.78513	3.38927	-1.25000	-0.86603	-0.85820	-0.86525
26	1.60000	2.30940	2.79550	2.57449	-1.00000	-0.69282	-0.63042	-0.66551
27	1.80000	2.59808	2.94721	3.13646	-0.75000	-0.51962	-0.45199	-0.51928
28	1.70000	2.45374	4.01387	3.15357	-0.50000	-0.34641	-0.29894	-0.32714
29	1.90000	2.74241	3.47887	3.27011	-0.25000	-0.17321	-0.17677	-0.16301
30	1.70000	2.45374	3.26431	2.89248	0.00000	0.00000	0.04783	0.03087
31	1.50000	2.16506	2.41301	2.57220	0.25000	0.17321	0.23779	0.16763
32	1.80000	2.59808	2.87394	3.12995	0.50000	0.34641	0.38646	0.36035
33	1.40000	2.02073	2.34406	2.22892	0.75000	0.51962	0.49893	0.48570
34	1.90000	2.74241	3.47064	3.61903	1.00000	0.69282	0.65094	0.66576
35	2.00000	2.88675	4.20559	3.28752	1.25000	0.86603	0.79511	0.86083
36	1.60000	2.30940	2.49622	2.58101	1.50000	1.03923	1.02635	1.03417
37	1.70000	2.45374	2.42736	2.72468	1.75000	1.21244	1.22281	1.23107
38	1.40000	2.02073	1.99253	2.27881	2.00000	1.38564	1.39037	1.34037
39	1.90000	2.74241	2.94510	3.18444	2.25000	1.55885	1.48530	1.52314
40	1.60000	2.30940	3.99991	3.23204	2.50000	1.73205	1.52512	1.63091
41	1.50000	2.16506	2.40032	2.03234	2.75000	1.90526	2.13695	2.01093
42	1.70000	2.45374	7.00000	6.04216	3.00000	2.07846	1.87981	1.87873
43	1.80000	2.59808	2.70006	1.73892	3.25000	2.25167	2.22019	2.79913
44	2.00000	2.88675	---	---	3.50000	2.42487	---	---
45	1.40000	2.02073	---	4.00385	3.75000	2.59808	---	2.31592

TABLE 9-7

Theoretical and Estimated Item Parameters in the 500 and 2,000 Subject Cases for Each Item of the Ten Item Test, Thirty-Five Item Test and the Additional Set of Thirty-Five Items. Three- Parameter Logistic Model Is Assumed. Case 4.

Item	Discrimination Parameter				Difficulty Parameter				Guessing Parameter	
	Theoretical		Estimated		Theoretical		Estimated		Estimated	
	Org.	Adj.	500 S.C.	2,000 S.C.	Org.	Adj.	500 S.C.	2,000 S.C.	500 S.C.	2,000 S.C.
1	1.50000	2.16506	3.41434	3.00167	-2.50000	-1.73205	-1.72520	-1.69465	0.00000	0.00000
2	1.00000	1.44338	3.29077	2.22717	-2.00000	-1.38564	-0.89904	-0.96868	0.40489	0.36573
3	2.50000	3.60844	4.26469	4.77358	-1.50000	-1.03923	-0.96121	-0.97525	0.04764	0.06512
4	1.00000	1.44338	1.49235	1.50609	-1.00000	-0.69282	-0.67263	-0.69374	0.02109	0.00572
5	1.50000	2.16506	2.69878	2.44557	-0.50000	-0.34641	-0.30645	-0.33458	0.00000	0.00473
6	1.00000	1.44338	1.45554	1.52362	0.00000	0.00000	-0.01051	-0.00980	0.00570	0.00000
7	2.00000	2.88675	2.81255	3.17229	0.50000	0.34641	0.35161	0.36437	0.00000	0.00000
8	1.00000	1.44338	1.61952	1.62062	1.00000	0.69282	0.66487	0.70467	0.00000	0.00000
9	2.00000	2.88675	3.88949	3.20578	1.50000	1.03923	1.00246	1.00496	0.00000	0.00000
10	1.00000	1.44338	1.49731	1.64327	2.00000	1.38564	1.35617	1.37264	0.00000	0.00000
11	1.80000	2.59808	---	---	-4.75000	-3.29090	---	---	---	---
12	1.90000	2.74241	---	---	-4.50000	-3.11769	---	---	---	---
13	2.00000	2.88675	---	---	-4.25000	-2.94449	---	---	---	---
14	1.50000	2.16506	7.00000	3.91266	-4.00000	-2.77128	-2.20782	-2.39100	0.00000	0.03468
15	1.60000	2.30940	---	2.18301	-3.75000	-2.59808	---	-3.12545	---	0.03468
16	1.40000	2.02073	7.00000	2.33617	-3.50000	-2.42487	-2.03372	-2.48901	0.50000	0.03468
17	1.90000	2.74241	3.35799	7.00000	-3.00000	-2.07846	-2.18794	-1.84693	0.11111	0.21436
18	1.80000	2.59808	2.25841	2.48088	-3.00000	-2.07846	-2.22797	-2.19957	0.11111	0.03468
19	1.60000	2.30940	6.56632	2.61144	-2.75000	-1.90526	-1.51892	-1.88161	0.50000	0.03468
20	2.00000	2.88675	4.32458	3.20291	-2.50000	-1.73205	-1.71620	-1.74176	0.00000	0.00000
21	1.50000	2.16506	3.10876	3.26381	-2.25000	-1.55885	-1.53891	-1.49560	0.00000	0.00000
22	1.70000	2.45374	2.61890	3.41314	-2.00000	-1.38564	-1.42058	-1.26670	0.00000	0.12604
23	1.50000	2.16506	2.82639	3.17517	-1.75000	-1.21244	-1.07270	-1.08862	0.15690	0.09335
24	1.40000	2.02073	2.06636	2.37464	-1.50000	-1.03923	-0.95700	-0.94428	0.03176	0.08417
25	2.00000	2.88675	3.57829	3.13381	-1.25000	-0.86603	-0.85408	-0.86278	0.00000	0.00000
26	1.60000	2.30940	2.87736	2.89571	-1.00000	-0.69282	-0.62853	-0.61657	0.00000	0.03865
27	1.80000	2.59808	3.03750	2.96977	-0.75000	-0.51962	-0.45344	-0.51406	0.00000	0.00000
28	1.70000	2.45374	3.73453	2.97059	-0.50000	-0.34641	-0.30369	-0.32323	0.00000	0.00000
29	1.90000	2.74241	3.46935	3.24521	-0.25000	-0.17321	-0.18461	-0.16184	0.00000	0.00000
30	1.70000	2.45374	3.11959	2.89623	0.00000	0.00000	0.03454	0.02895	0.00000	0.00000
31	1.50000	2.16506	2.34512	2.50594	0.25000	0.17321	0.22502	0.16486	0.00000	0.00000
32	1.80000	2.59808	2.77677	3.00145	0.50000	0.34641	0.37223	0.35562	0.00000	0.00000
33	1.40000	2.02073	2.26261	2.15307	0.75000	0.51962	0.49043	0.48439	0.00000	0.00000
34	1.90000	2.74241	3.27880	3.32280	1.00000	0.69282	0.64569	0.66430	0.00000	0.00000
35	2.00000	2.88675	3.49907	3.02911	1.25000	0.86603	0.79919	0.86670	0.00000	0.00000

36	1.60000	2.30940	2.26400	2.42606	1.50000	1.03923	1.04124	1.04434	0.00000	0.00000
37	1.70000	2.45374	2.46372	2.65747	1.75000	1.21244	1.23076	1.24084	0.00000	0.00000
38	1.40000	2.02073	2.01719	2.18450	2.00000	1.38564	1.39610	1.35317	0.00000	0.00000
39	1.90000	2.74241	2.69649	2.83214	2.25000	1.55885	1.51261	1.54567	0.00000	0.00000
40	1.60000	2.30940	3.55183	3.36066	2.50000	1.73205	1.55239	1.62235	0.00000	0.00000
41	1.50000	2.16506	1.90045	2.09162	2.75000	1.90526	2.29982	1.98522	0.00000	0.00000
42	1.70000	2.45374	7.00000	6.49245	3.00000	2.07846	1.85903	1.83539	0.00000	0.00000
43	1.80000	2.59808	2.22398	1.89946	3.25000	2.25167	2.34661	2.66871	0.00000	0.00000
44	2.00000	2.88675	---	---	3.50000	2.42487	---	---	---	---
45	1.40000	2.02073	---	3.53540	3.75000	2.59808	---	2.31707	---	0.00000
46	1.80000	2.59808	---	---	-3.75000	-2.59808	---	---	---	---
47	1.90000	2.74241	---	---	-3.50000	-2.42487	---	---	---	---
48	2.00000	2.88675	1.65615	2.48879	-3.25000	-2.25167	-2.89607	-2.46680	0.11111	0.03468
49	1.50000	2.16506	3.93845	2.11207	-3.00000	-2.07846	-1.95551	-2.26065	0.11111	0.03468
50	1.60000	2.30940	2.16634	3.30824	-2.75000	-1.90526	-2.00139	-1.60750	0.11111	0.50000
51	1.40000	2.02073	2.13178	2.73921	-2.50000	-1.73205	-1.71598	-1.51070	0.00000	0.27038
52	1.90000	2.74241	2.94297	3.13930	-2.00000	-1.38564	-1.42593	-1.42137	0.00000	0.00000
53	1.80000	2.59808	2.99335	3.04155	-2.00000	-1.38564	-1.20128	-1.30253	0.16075	0.08424
54	1.60000	2.30940	2.51813	2.56022	-1.75000	-1.21244	-1.19173	-1.22079	0.00000	0.00000
55	2.00000	2.88675	3.52683	3.56459	-1.50000	-1.03923	-0.97971	-0.98982	0.00220	0.00000
56	1.50000	2.16506	2.15083	2.45531	-1.25000	-0.86603	-0.81208	-0.86009	0.01728	0.00000
57	1.70000	2.45374	3.16197	2.83349	-1.00000	-0.69282	-0.59935	-0.67167	0.01181	0.00000
58	1.50000	2.16506	2.56946	2.43801	-0.75000	-0.51962	-0.49076	-0.51983	0.00000	0.00000
59	1.40000	2.02073	2.47960	2.50390	-0.50000	-0.34641	-0.30105	-0.31697	0.00000	0.00000
60	2.00000	2.88675	2.94042	3.29908	-0.25000	-0.17321	-0.15217	-0.15679	0.00000	0.00000
61	1.60000	2.30940	2.92132	2.52572	0.00000	0.00000	-0.01023	0.02219	0.00000	0.00000
62	1.80000	2.59808	2.85132	2.99588	0.25000	0.17321	0.21551	0.20297	0.00000	0.00000
63	1.70000	2.45374	2.62639	2.74542	0.50000	0.34641	0.31140	0.33140	0.00000	0.00000
64	1.90000	2.74241	3.04783	3.06757	0.75000	0.51962	0.53653	0.54103	0.00000	0.00000
65	1.70000	2.45374	3.18272	2.89775	1.00000	0.69282	0.67249	0.66449	0.00000	0.00000
66	1.50000	2.16506	2.70700	2.48166	1.25000	0.86603	0.83533	0.85440	0.00000	0.00000
67	1.80000	2.59808	3.02039	3.41981	1.50000	1.03923	0.96838	1.02375	0.00000	0.00000
68	1.40000	2.02073	2.34530	2.36776	1.75000	1.21244	1.25342	1.16936	0.00000	0.00000
69	1.90000	2.74241	2.86873	3.24824	2.00000	1.38564	1.34379	1.35137	0.00000	0.00000
70	2.00000	2.88675	3.04314	2.87183	2.25000	1.55885	1.48181	1.52512	0.00000	0.00000
71	1.60000	2.30940	4.58205	3.17071	2.50000	1.73205	1.60055	1.66385	0.00000	0.00000
72	1.70000	2.45374	2.87135	3.68760	2.75000	1.90526	1.87762	1.83679	0.00000	0.00000
73	1.40000	2.02073	7.00000	2.70660	3.00000	2.07846	1.92387	2.05018	0.01188	0.00000
74	1.90000	2.74241	---	7.00000	3.25000	2.25167	---	2.06124	---	0.00000
75	1.60000	2.30940	2.95348	3.20156	3.50000	2.42487	2.45972	2.33965	0.00000	0.00000
76	1.50000	2.16506	---	3.62677	3.75000	2.59808	---	2.53398	---	0.00000
77	1.70000	2.45374	7.00000	6.87045	4.00000	2.77128	2.09473	2.14071	0.00000	0.00000
78	1.80000	2.59808	---	---	4.25000	2.94449	---	---	---	---
79	2.00000	2.88675	---	---	4.50000	3.11769	---	---	---	---
80	1.40000	2.02073	---	---	4.75000	3.29090	---	---	---	---

TABLE 9-8

Theoretical and Estimated Item Parameters in the 500 and 2,000 Subject Cases for Each Item of the Ten Item Test, Thirty-Five Item Test and the Additional Set of Thirty-Five Items. Logistic Model Is Assumed. Case 4.

Item	Discrimination Parameter				Difficulty Parameter			
	Theoretical		Estimated		Theoretical		Estimated	
	Org.	Adj.	500 S.C.	2,000 S.C.	Org.	Adj.	500 S.C.	2,000 S.C.
1	1.50000	2.16506	3.72684	3.07028	-2.50000	-1.73205	-1.69210	-1.68238
2	1.00000	1.44338	1.65223	1.52731	-2.00000	-1.38564	-1.36793	-1.37154
3	2.50000	3.60844	3.59656	3.70406	-1.50000	-1.03923	-1.01728	-1.03414
4	1.00000	1.44338	1.44195	1.49512	-1.00000	-0.69282	-0.70755	-0.70424
5	1.50000	2.16506	2.67224	2.38228	-0.50000	-0.34641	-0.31213	-0.34367
6	1.00000	1.44338	1.41840	1.51732	0.00000	0.00000	-0.02249	-0.01020
7	2.00000	2.88675	2.78154	3.15697	0.50000	0.34641	0.35386	0.36679
8	1.00000	1.44338	1.60933	1.62039	1.00000	0.69282	0.66845	0.70644
9	2.00000	2.88675	3.86351	3.22524	1.50000	1.03923	1.00912	1.00685
10	1.00000	1.44338	1.50703	1.65231	2.00000	1.38564	1.35715	1.37203
11	1.80000	2.59808	---	---	-4.75000	-3.29090	---	---
12	1.90000	2.74241	---	---	-4.50000	-3.11769	---	---
13	2.00000	2.88675	---	---	-4.25000	-2.94449	---	---
14	1.50000	2.16506	7.00000	4.18964	-4.00000	-2.77128	-2.12460	-2.30785
15	1.60000	2.30940	---	2.26449	-3.75000	-2.59808	---	-3.05746
16	1.40000	2.02073	4.93831	2.48782	-3.50000	-2.42487	-2.21029	-2.42230
17	1.90000	2.74241	3.80449	7.00000	-3.00000	-2.07846	-2.12493	-1.85373
18	1.80000	2.59808	2.44437	2.62452	-3.00000	-2.07846	-2.18906	-2.15717
19	1.60000	2.30940	3.05923	2.65246	-2.75000	-1.90526	-1.80634	-1.87546
20	2.00000	2.88675	4.53018	3.38488	-2.50000	-1.73205	-1.68883	-1.71972
21	1.50000	2.16506	3.21950	3.34875	-2.25000	-1.55885	-1.52729	-1.49074
22	1.70000	2.45374	2.75628	3.01913	-2.00000	-1.38564	-1.41352	-1.34519
23	1.50000	2.16506	2.36370	2.77310	-1.75000	-1.21244	-1.21285	-1.16388
24	1.40000	2.02073	2.00406	2.11082	-1.50000	-1.03923	-0.99610	-1.02829
25	2.00000	2.88675	3.57138	3.13369	-1.25000	-0.86603	-0.86755	-0.86954
26	1.60000	2.30940	2.84127	2.51903	-1.00000	-0.69282	-0.63886	-0.66645
27	1.80000	2.59808	2.99887	2.95792	-0.75000	-0.51962	-0.46131	-0.51804
28	1.70000	2.45374	3.66463	2.95891	-0.50000	-0.34641	-0.30923	-0.32549
29	1.90000	2.74241	3.41412	3.23105	-0.25000	-0.17321	-0.18852	-0.16273
30	1.70000	2.45374	3.09273	2.88274	0.00000	0.00000	0.03311	0.02941
31	1.50000	2.16506	2.31939	2.49396	0.25000	0.17321	0.22567	0.16600
32	1.80000	2.59808	2.74806	2.99118	0.50000	0.34641	0.37470	0.35793
33	1.40000	2.02073	2.24698	2.15137	0.75000	0.51962	0.49371	0.48654
34	1.90000	2.74241	3.25685	3.32014	1.00000	0.69282	0.65121	0.66740
35	2.00000	2.88675	3.46982	3.04180	1.25000	0.86603	0.80565	0.86921

36	1.60000	2.30940	2.26525	2.43695	1.50000	1.03923	1.04618	1.04584
37	1.70000	2.45374	2.49287	2.67665	1.75000	1.21244	1.23358	1.24116
38	1.40000	2.02073	2.03327	2.19706	2.00000	1.38564	1.39735	1.35292
39	1.90000	2.74241	2.73122	2.85681	2.25000	1.55885	1.51178	1.54391
40	1.60000	2.30940	3.60493	3.36654	2.50000	1.73205	1.55051	1.62142
41	1.50000	2.16506	1.96429	2.09506	2.75000	1.90526	2.27272	1.98431
42	1.70000	2.45374	7.00000	6.25673	3.00000	2.07846	1.85406	1.83992
43	1.80000	2.59808	2.24123	1.89216	3.25000	2.25167	2.33998	2.67372
44	2.00000	2.88675	---	---	3.50000	2.42487	---	---
45	1.40000	2.02073	---	3.44390	3.75000	2.59808	---	2.33500
46	1.80000	2.59808	---	---	-3.75000	-2.59808	---	---
47	1.90000	2.74241	---	---	-3.50000	-2.42487	---	---
48	2.00000	2.88675	1.65181	2.77341	-3.25000	-2.25167	-2.92148	-2.36388
49	1.50000	2.16506	4.29085	2.19098	-3.00000	-2.07846	-1.92503	-2.23043
50	1.60000	2.30940	2.15004	2.64628	-2.75000	-1.90526	-2.02814	-1.87139
51	1.40000	2.02073	2.26422	2.46143	-2.50000	-1.73205	-1.68791	-1.65402
52	1.90000	2.74241	3.11145	3.20821	-2.00000	-1.38564	-1.41906	-1.41957
53	1.80000	2.59808	2.53858	2.88308	-2.00000	-1.38564	-1.32469	-1.35346
54	1.60000	2.30940	2.56117	2.57874	-1.75000	-1.21244	-1.19716	-1.22428
55	2.00000	2.88675	3.53579	3.56475	-1.50000	-1.03923	-0.99484	-0.99738
56	1.50000	2.16506	2.08227	2.45651	-1.25000	-0.86603	-0.84039	-0.86580
57	1.70000	2.45374	2.93475	2.81444	-1.00000	-0.69282	-0.62403	-0.67700
58	1.50000	2.16506	2.55209	2.42151	-0.75000	-0.51962	-0.49880	-0.52372
59	1.40000	2.02073	2.45314	2.48398	-0.50000	-0.34641	-0.30666	-0.31923
60	2.00000	2.88675	2.92942	3.28162	-0.25000	-0.17321	-0.15584	-0.15762
61	1.60000	2.30940	2.89170	2.50830	0.00000	0.00000	-0.01218	0.02247
62	1.80000	2.59808	2.82379	2.98334	0.25000	0.17321	0.21617	0.20454
63	1.70000	2.45374	2.59042	2.72777	0.50000	0.34641	0.31316	0.33348
64	1.90000	2.74241	3.01309	3.06215	0.75000	0.51962	0.54090	0.54394
65	1.70000	2.45374	3.15705	2.90078	1.00000	0.69282	0.67811	0.66738
66	1.50000	2.16506	2.69905	2.48647	1.25000	0.86603	0.84115	0.85672
67	1.80000	2.59808	3.06261	3.43685	1.50000	1.03923	0.97402	1.02562
68	1.40000	2.02073	2.38477	2.37751	1.75000	1.21244	1.25516	1.17030
69	1.90000	2.74241	2.87444	3.25895	2.00000	1.38564	1.34676	1.35142
70	2.00000	2.88675	3.07327	2.88669	2.25000	1.55885	1.48189	1.52410
71	1.60000	2.30940	4.65176	3.18260	2.50000	1.73205	1.59754	1.66249
72	1.70000	2.45374	2.93033	3.68052	2.75000	1.90526	1.86943	1.83665
73	1.40000	2.02073	3.03207	2.69563	3.00000	2.07846	2.05377	2.05217
74	1.90000	2.74241	---	7.00000	3.25000	2.25167	---	2.06555
75	1.60000	2.30940	2.97866	3.09922	3.50000	2.42487	2.45256	2.36309
76	1.50000	2.16506	---	3.44436	3.75000	2.59808	---	2.57899
77	1.70000	2.45374	7.00000	6.45131	4.00000	2.77128	2.09808	2.16256
78	1.80000	2.59808	---	---	4.25000	2.94449	---	---
79	2.00000	2.88675	---	---	4.50000	3.11769	---	---
80	1.40000	2.02073	---	---	4.75000	3.29090	---	---

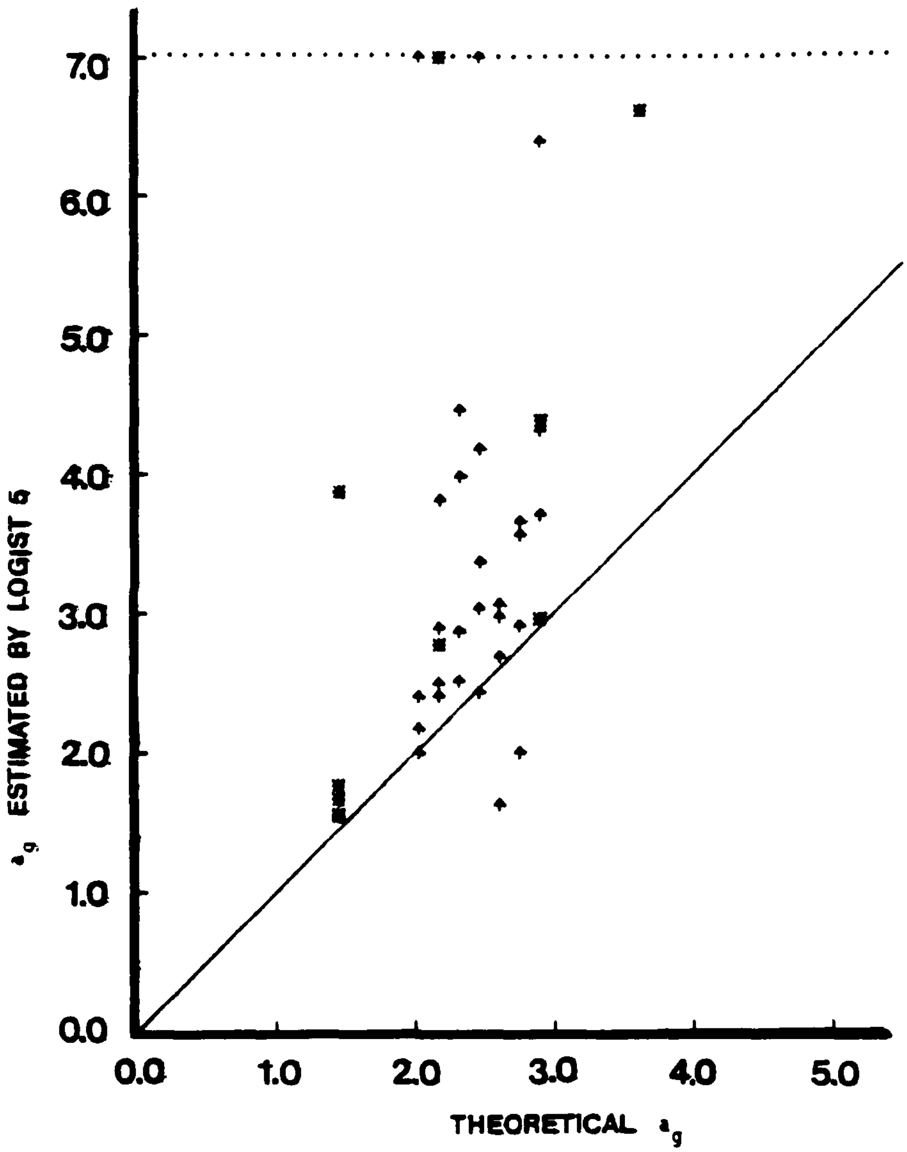


FIGURE 9-1

Estimated Discrimination Parameter Obtained by Logist 5 Plotted against the True Discrimination Parameter a_g for Each Item of the Ten Item Test (✱) and of the Thirty-Five Item Test (♦). Case 3, 500 Subject Case. Three-Parameter Logistic Model Is Assumed.

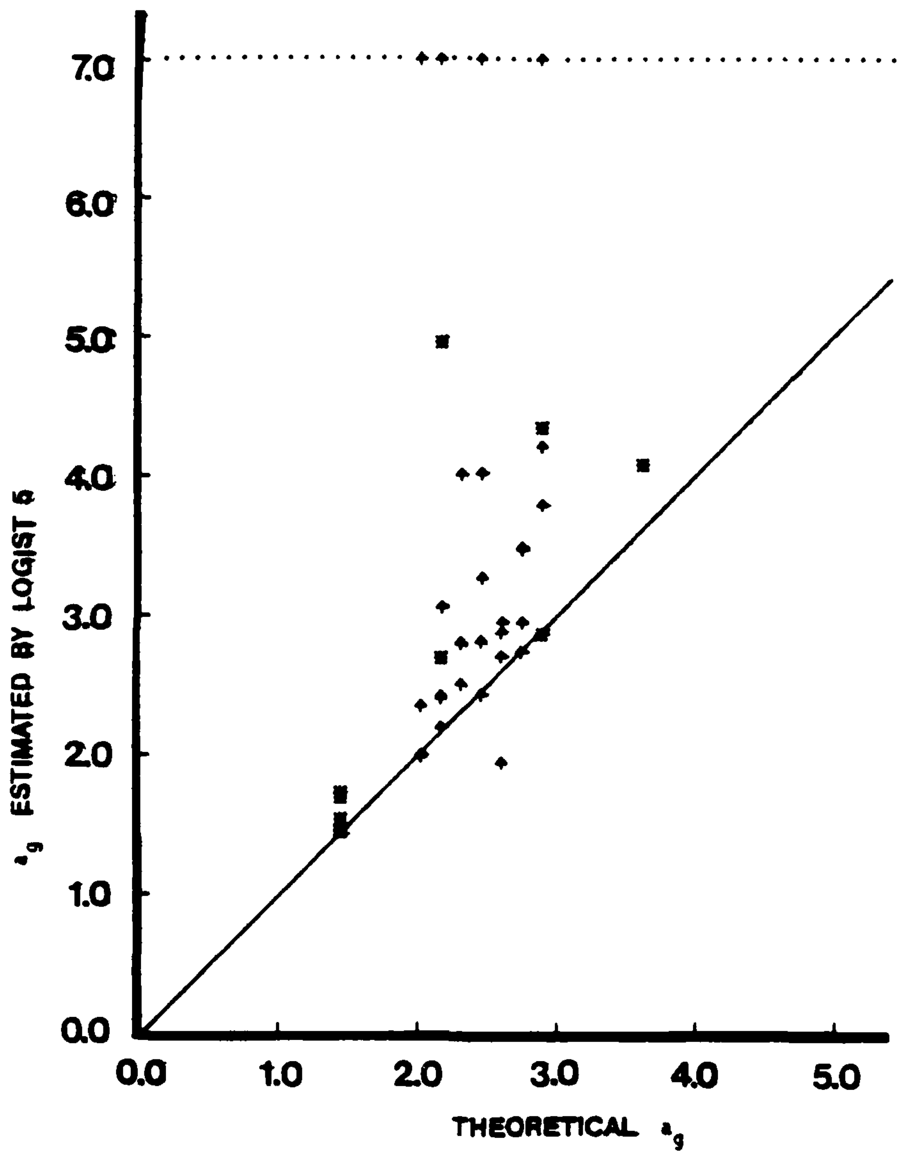


FIGURE 9-1 (Continued)

Guessing Parameter c_g Is Set Equal to Zero, i.e., Logistic Model Is Assumed.

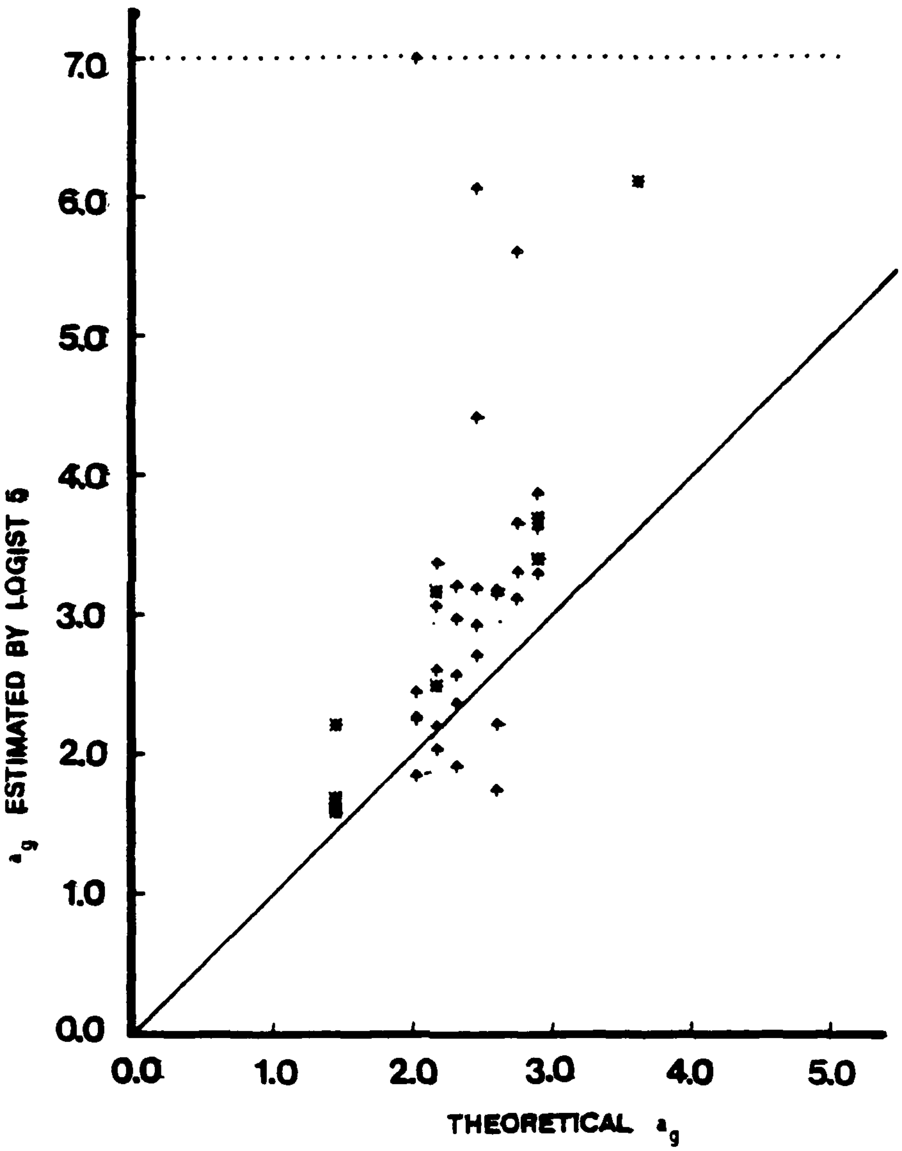


FIGURE 9-2

Estimated Discrimination Parameter Obtained by Logist 5 Plotted against the True Discrimination Parameter a_g for Each Item of the Ten Item Test (✱) and of the Thirty-Five Item Test (✦). Case 3, 2,000 Subject Case. *Three-Parameter Logistic Model Is Assumed.*

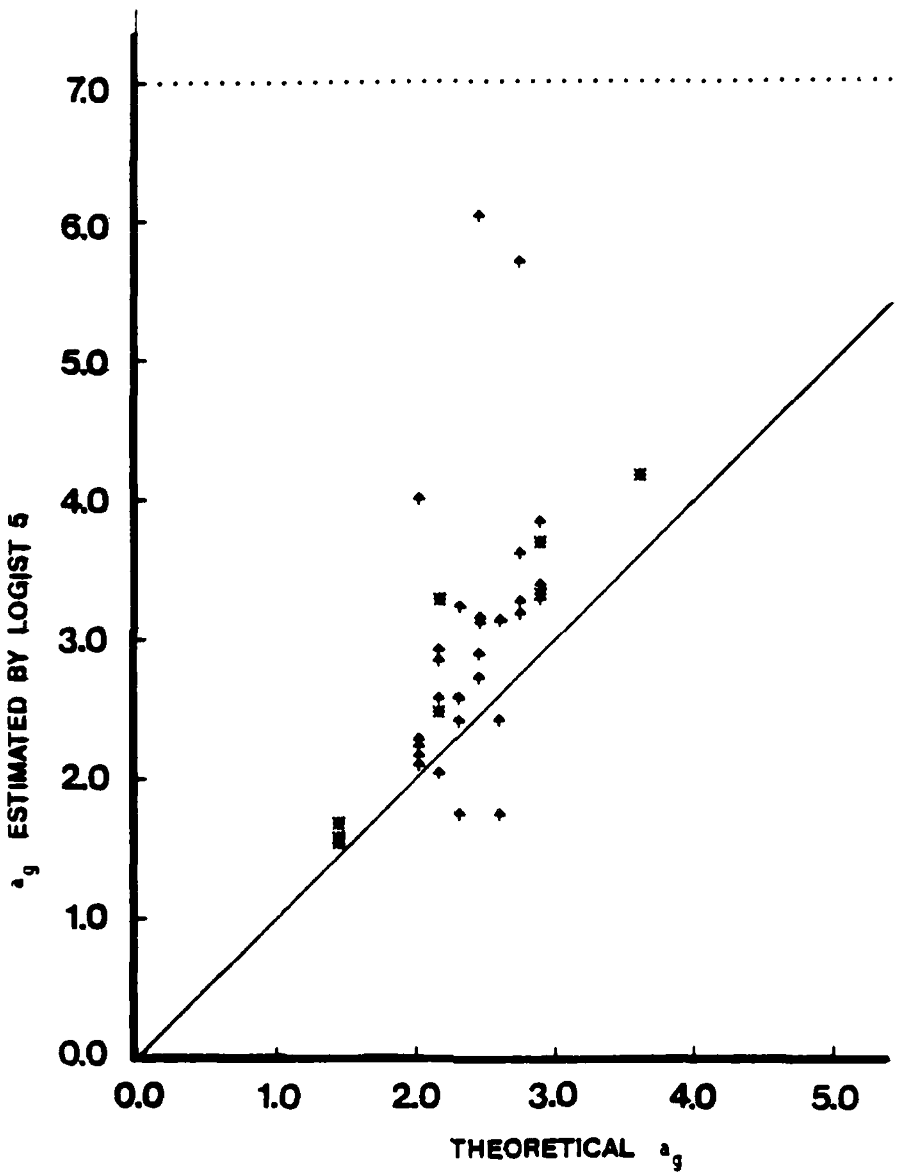


FIGURE 9-2 (Continued)

Guessing Parameter c_g Is Set Equal to Zero, i.e., Logistic Model Is Assumed.

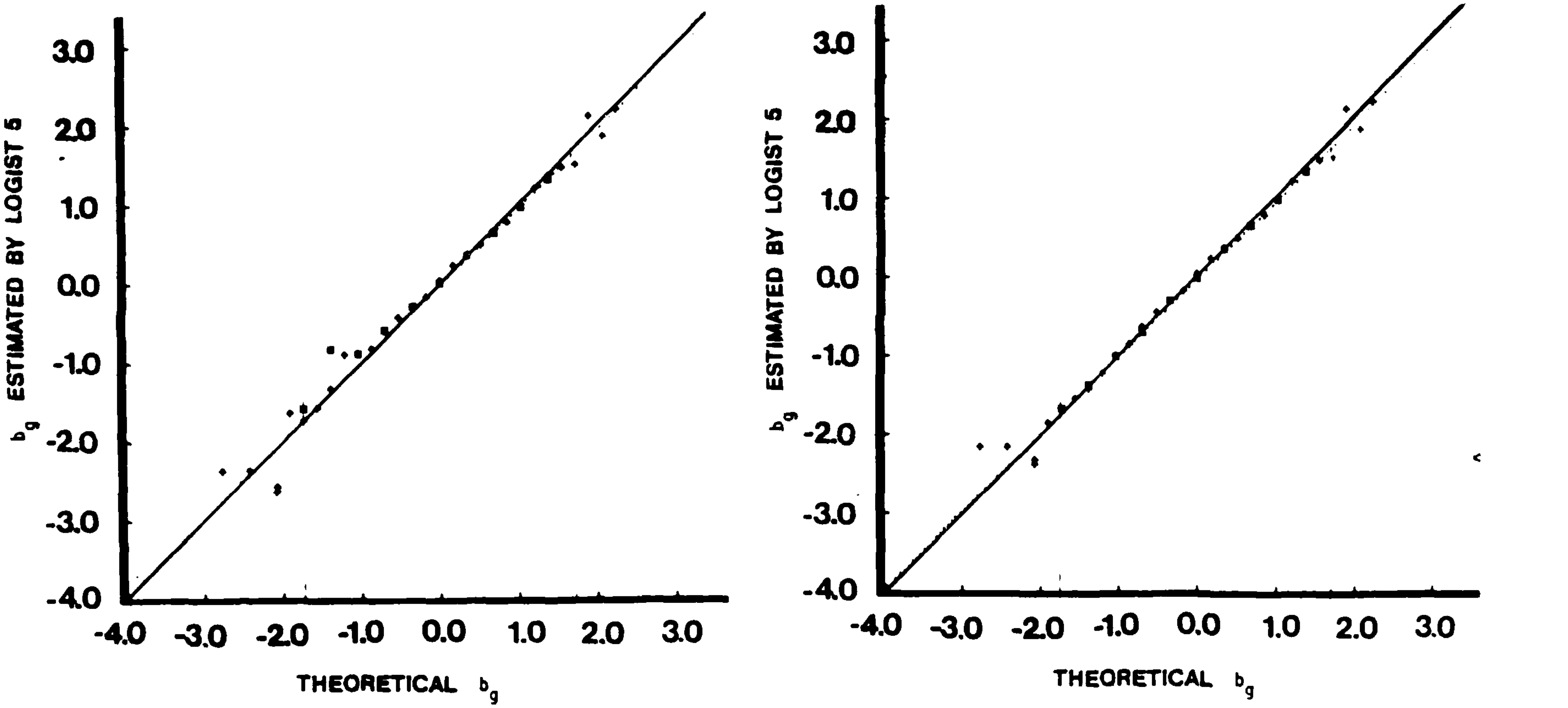


FIGURE 9-3

Estimated Difficulty Parameter Obtained by Logist 5 Plotted against the True Difficulty Parameter b_g for Each Item of the Ten Item Test (*) and of the Thirty-Five Item Test (♦). Case 3, 500 Subject Case. Three-Parameter Logistic Model Is Assumed (Left). Guessing Parameter c_g Is Set Equal to Zero, i.e., Logistic Model Is Assumed (Right). Linear Regression of $\hat{\theta}$ on θ Is Plotted by Dots for Reference.

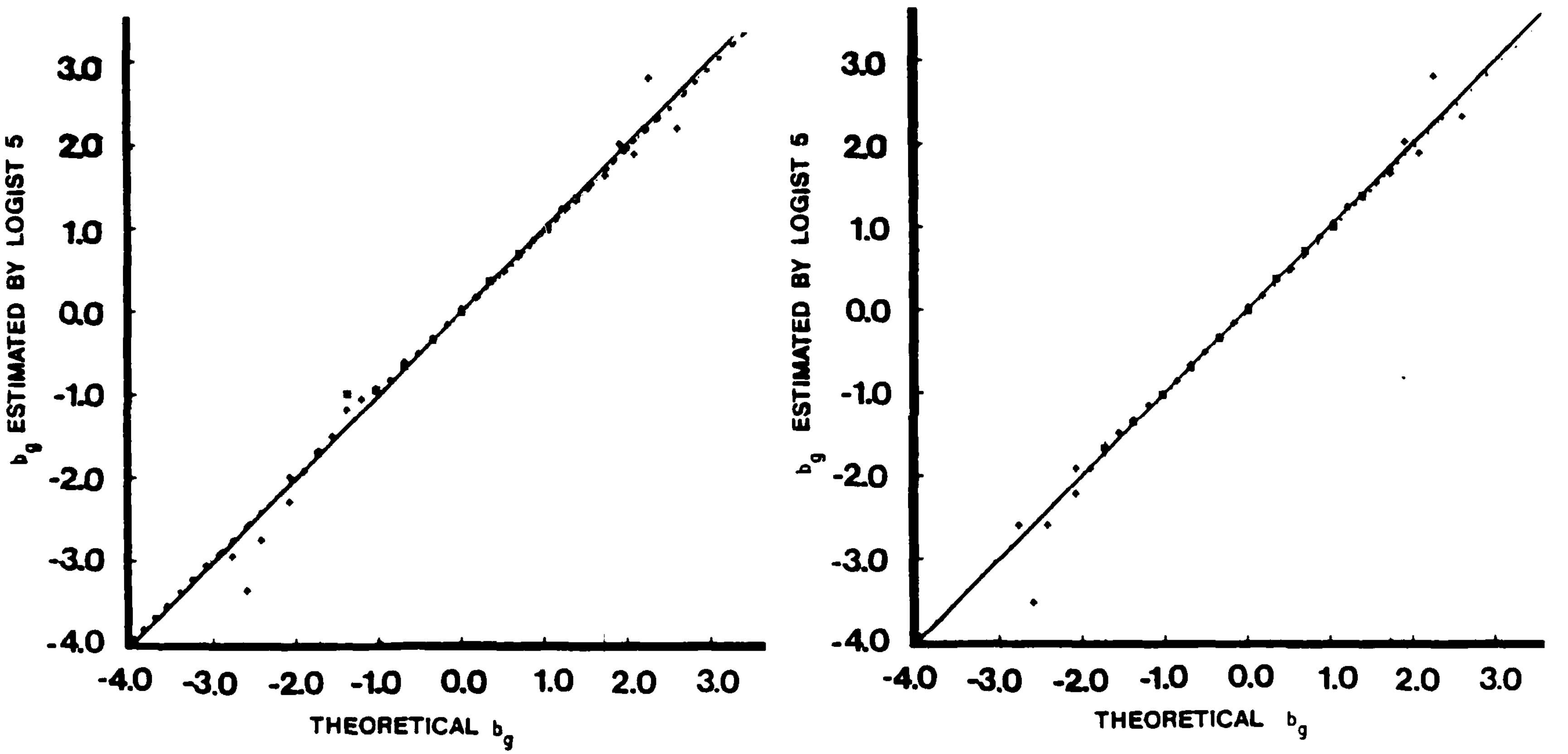


FIGURE 9-4

Estimated Difficulty Parameter Obtained by Logist 5 Plotted against the True Difficulty Parameter b_g for Each Item of the Ten Item Test (*) and of the Thirty-Five Item Test (♦). Case 3, 2,000 Subject Case. Three-Parameter Logistic Model is Assumed (Left). Guessing Parameter c_g Is Set Equal to Zero, i.e., Logistic Model Is Assumed (Right). Linear Regression of $\hat{\theta}$ on θ Is Plotted by Dots for Reference.

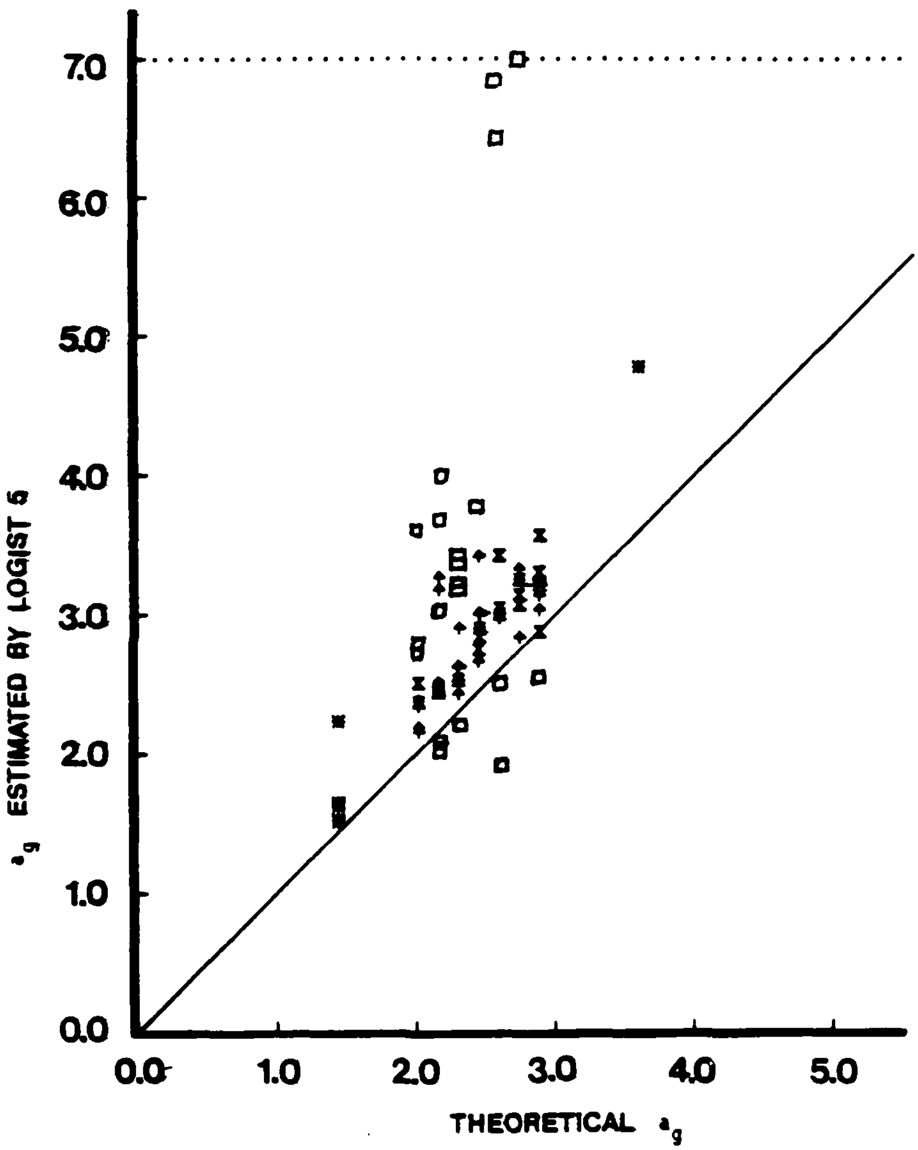


FIGURE 9-5

Reduced Scatter Diagram of the Theoretical and Estimated Item Discrimination Parameters Obtained by Excluding All Items Whose Theoretical Item Difficulty Parameters Are Outside the Interval $(-\sqrt{3}, \sqrt{3})$. These Excluded Items Are Plotted by Hollow Shapes. Case 4, 2,000 Subject Case.

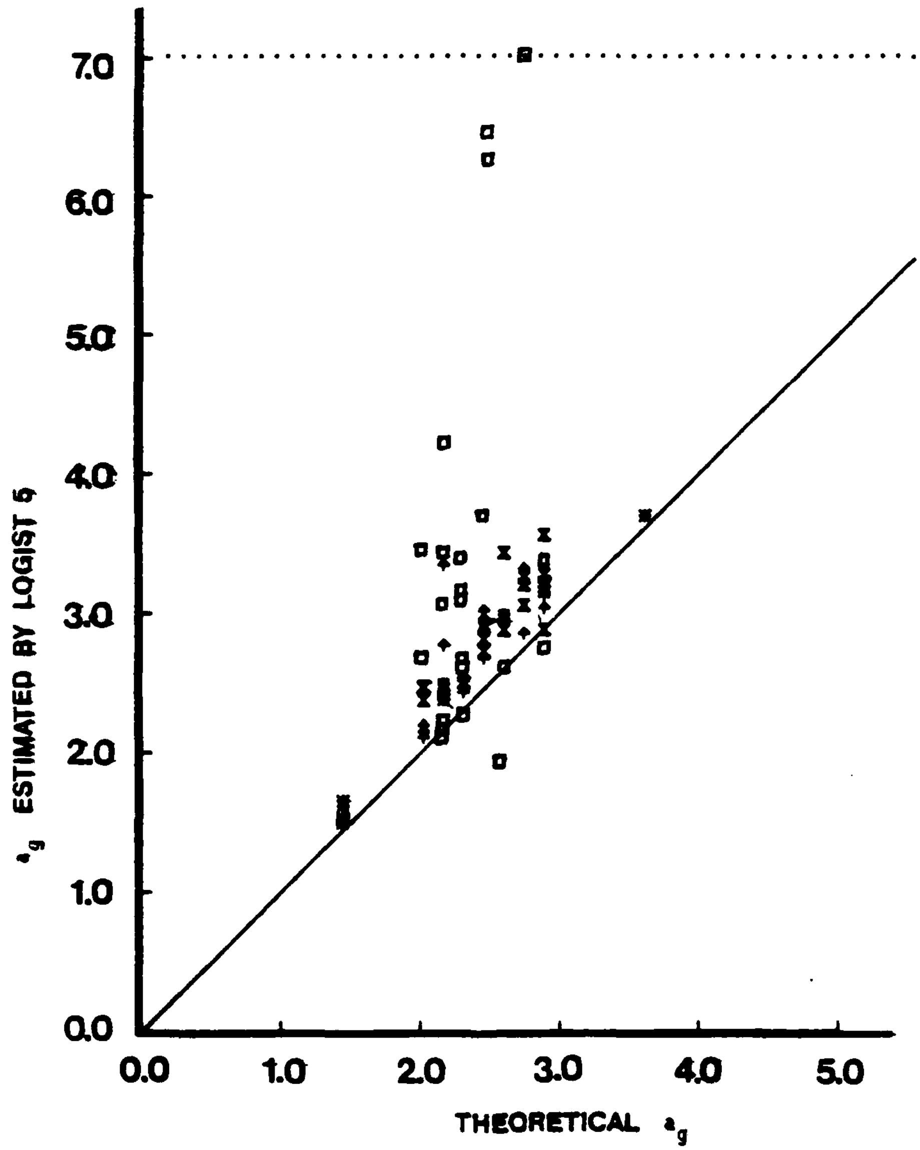


FIGURE 9-5 (Continued)

Guessing Parameter c_g Is Set Equal to Zero, i.e., Logistic Model Is Assumed.

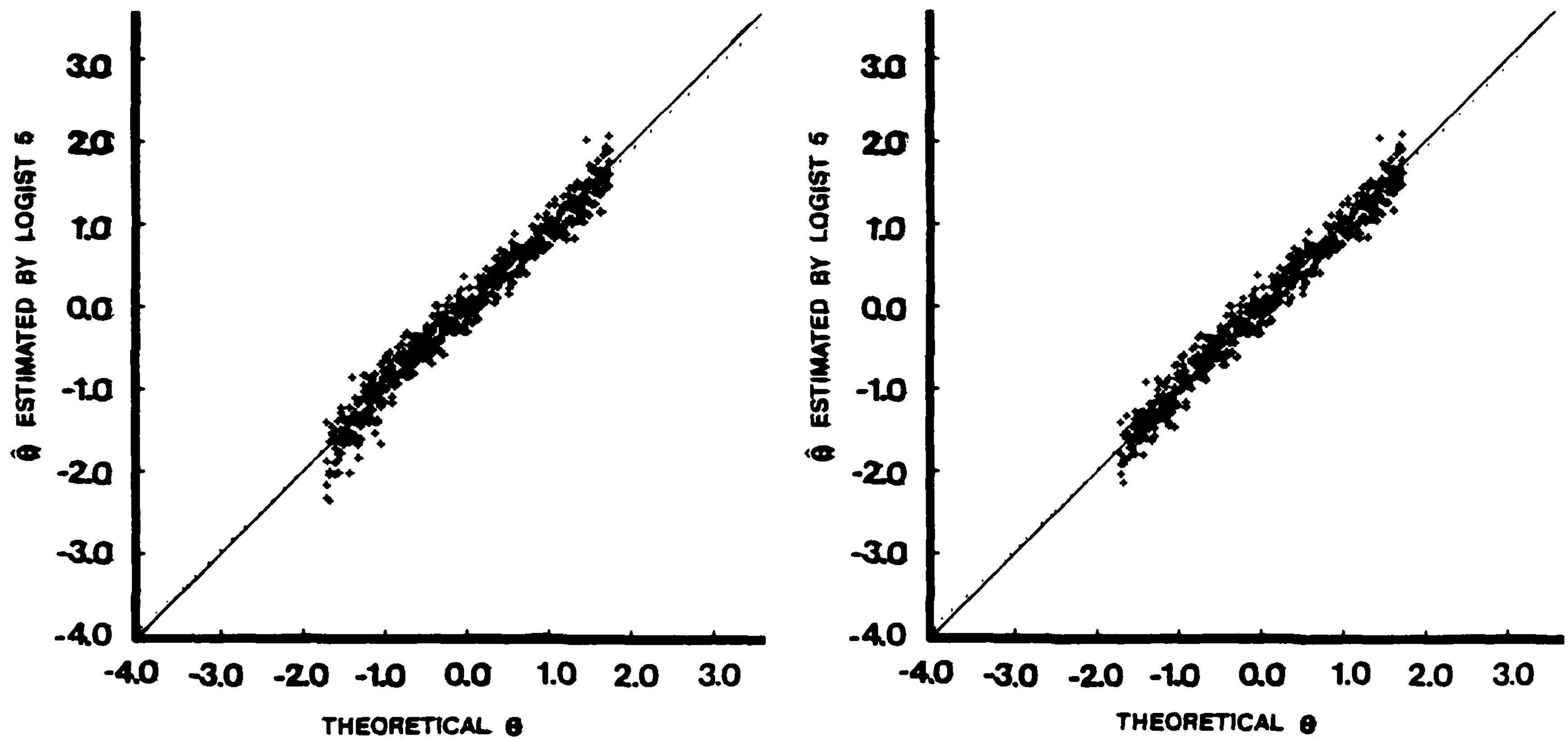


FIGURE 9-6

Estimated Individual Parameters Plotted against the Theoretical Individual Parameters. Three-parameter Logistic Model (Left Graph) and Logistic Model (Right Graph) Are Assumed, Respectively. Case 3, 500 Subject Case. Linear Regression of $\hat{\theta}$ on θ Is Plotted by a Dotted Line.

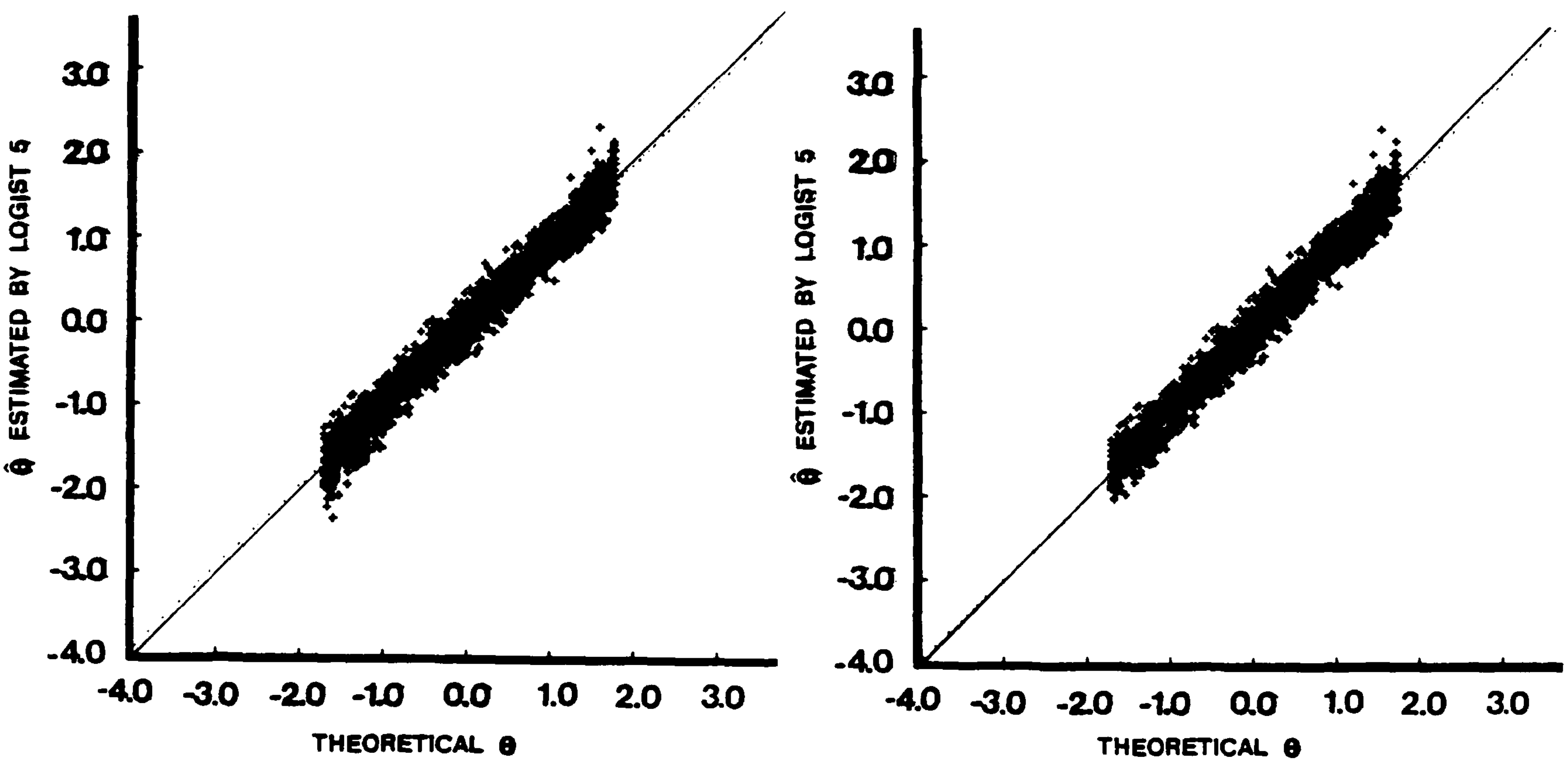


FIGURE 9-7

Estimated Individual Parameters Plotted against the Theoretical Individual Parameters. Three-parameter Logistic Model (Left Graph) and Logistic Model (Right Graph) Are Assumed, Respectively. Case 3, 2,000 Subject Case. Linear Regression of $\hat{\theta}$ on θ Is Plotted by a Dotted Line.

which the ability distribution is uniform, is indicated by two solid, vertical lines. A reasonably good agreement is observed between the estimated and the theoretical difficulty parameters for the subset of items whose theoretical difficulty parameters are within the interval, $(-\sqrt{3}, \sqrt{3})$. Some improvement is observed in the results obtained by assuming the logistic model compared with those obtained by assuming the three-parameter logistic model, both in the 500 Subject Case and in the 2,000 Case, in each of the Cases 1, 2, 3 and 4.

Figure 9-5 presents the estimated item discrimination parameters of Case 4 plotted against the true discrimination parameters. In this figure, all items whose difficulty parameters are outside the range of $(-\sqrt{3}, \sqrt{3})$ are plotted by hollow shapes. Case 4 has produced the best agreement of the four cases in each situation, and we can see that agreement is further improved by excluding these hollow shapes. A substantial enhancement of the estimated discrimination parameters still exists, however. The resulting estimated item characteristic functions are all compared to the theoretical curves in graphs in [I.2.8].

Figures 9-6 and 9-7 illustrate the estimated individual parameters $\hat{\theta}_s$ plotted against the true values θ_s for the 500 and 2,000 Subject Cases, respectively, in Case 3 where we have thirty-test items.

[IX.4] Discriminating Shrinkage Factor and Difficulty Reduction Index

It has been observed that the enhancement of the estimated discrimination parameter exists in both situations where we set the third parameter c_g free and at zero in Logist 5, respectively. This indicates the effect of scaling problem in Logist 5 where the standard deviation of the maximum likelihood estimate of $\hat{\theta}_s$, instead of that of θ_s , is defined as the unit. Since the standard deviation of $\hat{\theta}_s$ is expected to be larger than θ_s for the kind of data such as ours, the estimated discrimination parameters are expected to be enhanced, and the estimated difficulty parameters are expected to be regressed, than what they actually are.

It has also been observed that the enhancement of the estimated discrimination parameter tends to be larger in the situation where three-parameter logistic model is assumed, in comparison with the situation where c_g is set equal to zero. This fact suggests that there exists the enhancement of the estimated discrimination parameter caused by the third parameter c_g . It is also suggested from the estimated difficulty parameters that the enhancement of the estimated difficulty parameter will also exist when three-parameter logistic model is assumed, if an appropriate scale adjustment of θ is made.

For these reasons, the principal investigator proposed the discrimination shrinkage factor and the difficulty reduction index, following certain rationale. They are given by

$$(9.3) \qquad [\varsigma(c_g^*)]^{-1} = [\log(1-c_g^*) - \log(1+c_g^*)][\log(1-2c_g^*]^{-1}$$

and

$$(9.4) \qquad \xi(c_g^* \mid a_g) = (Da_g)^{-1}[\log(1+c_g^*) - \log(1-c_g^*)] \ ,$$

where c_g is the estimated positive third parameter when it actually should equal zero. Thus we can revise the estimated discrimination parameter a_g^* by setting

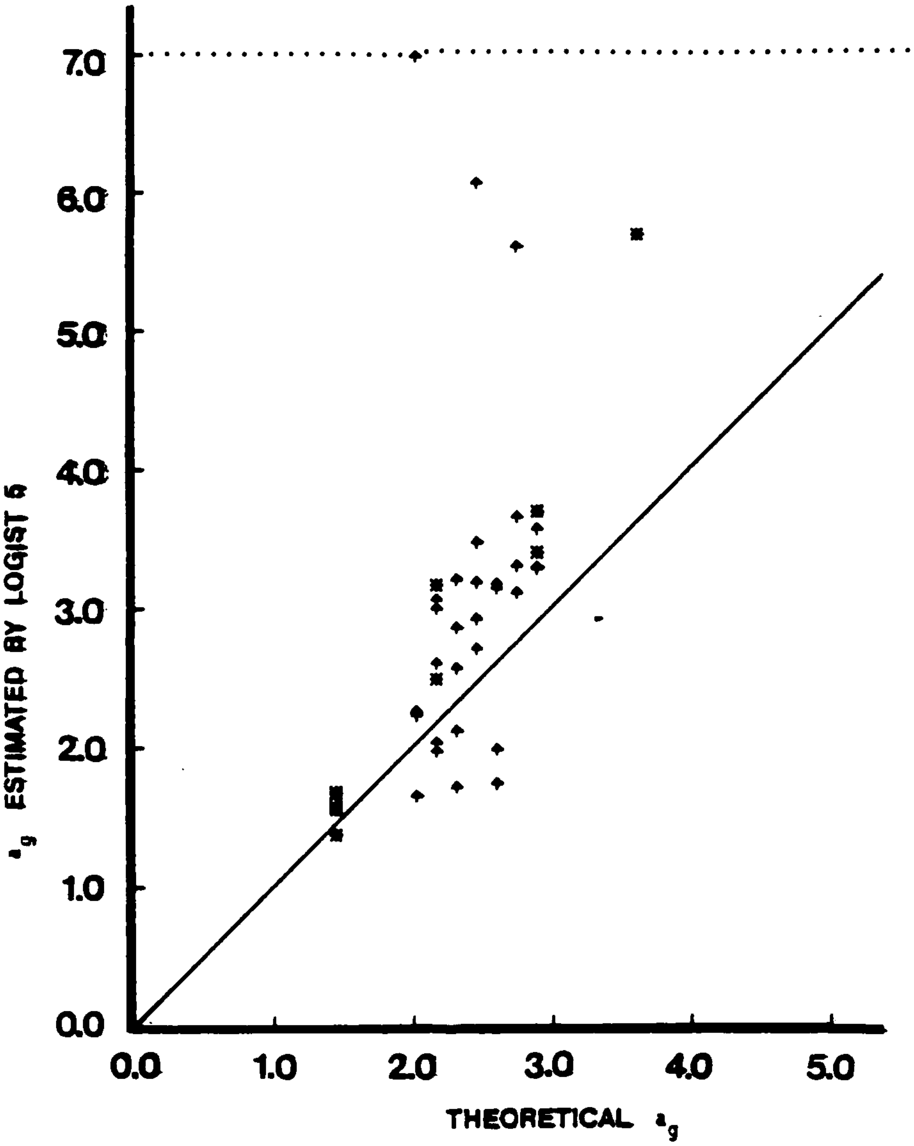


FIGURE 9-8

Estimated Discrimination Parameters Which Were Revised by the Discrimination Shrinkage Factor Plotted against the Theoretical Discrimination Parameters. Three-Parameter Logistic Model Is Assumed. Case 3, 2,000 Subject Case.

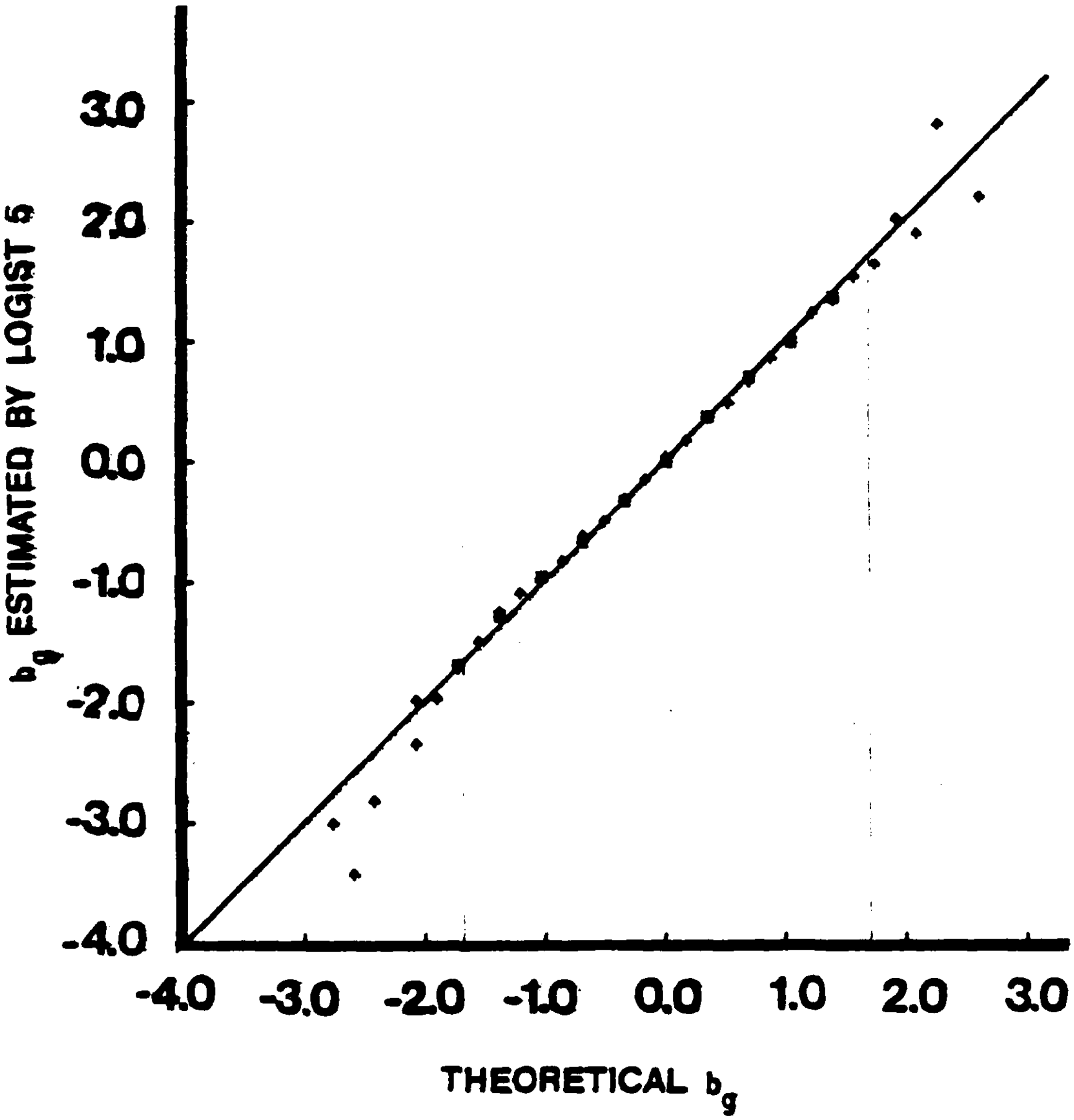


FIGURE 9-9

Estimated Difficulty Parameters Which Were Revised by the Difficulty Reduction Index Plotted against the Theoretical Difficulty Parameters. Three-Parameter Logistic Model Is Assumed. Case 3, 2,000 Subject Case.

$$\hat{a}_g = a_g^*[\varsigma(c_g^*)] \quad , \tag{9.5}$$

and the estimated difficulty parameter b_g^* by

$$\hat{b}_g = b_g^* - \varsigma(c_g^* \mid \hat{a}_g) \quad . \tag{9.6}$$

Figures 9-8 and 9-9 illustrate the results of these revisions for the 2,000 Subject Case in Case 3. Comparison of these figures with Figures 9-2 and 9-4 indicates the effects of these revisions.

[IX.5] Discussion

The present research disclosed the danger of accepting the results of Logist 5 without modifications. It is the principal investigator's wish that researchers become aware of the danger, and think twice before accepting the results as they are.

Again in this chapter only a small part of the research was presented because of shortage of space. There are many more interesting findings and observations contained in [I.1.8].

X Discussion and Conclusions

In writing this final report, the principal investigator has found it extremely difficult to summarize and integrate all the different approaches and findings of the research. They include: (1) basically theoretical works such as the proposal of two new latent trait models (cf. Chapters V and VI), that of the MLE bias function for general discrete responses (cf. Chapter III), constancy in item information and the information loss caused by noise (cf. Chapter IV), etc., 2) combinations of theoretical and methodological works including a substantial amount of computer programming such as the estimation of the operating characteristics of discrete item responses (cf. Chapter II), etc., 3) basically empirical studies such as the analyses of the Iowa Data and that of Shiba's data (cf. Chapters VII and VIII), and 4) basically a simulation study like the observations of the results of Logist 5 (cf. Chapter IX). The above is a rough categorization, for the contents are overlapping. To give some examples, the proposal of the discrimination shrinkage factor and the difficulty reduction index in Chapter IX are theoretical works, the plausibility function of a distractor observed in Chapter VII has something to do with the Informative Distractor Model, and all theoretical works have some empirical studies or their prospects involved.

The principal investigator believes that all these different approaches are essential to the advancement of latent trait theory, and are the reason for the fruitfulness of the present research project. She regrets, however, that she had to leave out many other interesting findings and observations from this final report, because of the shortage of space and of the difficulty in summarizing all of them. She hopes that readers of this final report who have got interested in particular topics will read the original research reports and/or conference handouts.

The principal investigator also believes that she has accomplished something during this research period, in line with the proposed objectives for the advancement of latent trait theory. At the end of this research project, she would like to express her gratitude to the Office of Naval Research for this research opportunity.

APPENDIX A

Contents of Advanced Seminar on Latent Trait Theory (1982)

I Estimation of the Operating Characteristics of the Discrete Item Responses and That of Ability Distributions: I

(I.1) Relationship between the Estimation of the Operating Characteristics and That of Ability Distributions

(I.2) No Mathematical Forms are Assumed for the Operating Characteristics of the Unknown Test Items

(I.3) Small Number of Examinees in the Calibration Data

(I.4) Old Test

(I.5) Set of Five Hundred Maximum Likelihood Estimates

(I.6) Unknown Test Items Whose Operating Characteristics Are to Be Estimated

(I.7) Use of Robust, Indirect Information

(I.8) Transformation of Ability θ to τ

II Method of Moments As the Least Squares Solution for Fitting a Polynomial

(II.1) Approximation to the Density Function from a Set of Observations

(II.2) Method of Moments As the Least Squares Solution for Fitting a Polynomial

(II.3) Direct Use of the Least Squares Solution

(II.4) Solution by the Method of Moments

(II.5) Expanded Use of the Method of Moments

(II.6) Selection of the Interval

(II.7) Comparison of the Results Obtained by the Method of Moments and by the Direct Least Squares Procedure

III Estimation of the Operating Characteristics of the Discrete Item Responses and That of Ability Distributions: II

(III.1) Estimated Operating Characteristics Which Are Directly Observable from Our Calibration Data

(III.2) Necessary Correction for the Scale of the Maximum Likelihood Estimate When Used As a Substitute for the Ability Scale

(III.3) Transformation of θ to τ Using the Method of Moments for Fitting a Polynomial

(III.4) Classification of Methods and Approaches

(III.5) Normal Approximation Method

VI Constant Information Model

VII A New Family of Models for the Multiple-Choice Test Item: I

VIII New Family of Models for the Multiple-Choice Test Item: II

(VIII.1) Shiba's Word Comprehension Tests

(VIII.2) Subjects Used in Shiba's Research

(VIII.3) Methods and Results of Shiba's Research

(VIII.4) Distractors As Resources of Information

(VIII.5) Mathematical Models in Physics and in Psychology

(VIII.6) Normal Ogive Model on the Graded Response Level and Bock's Multinomial Model

(VIII.7) A New Family of Models for the Multiple-Choice Test Items

(VIII.8) Basic Functions and Information Functions of the New Models

(VIII.9) Instructions and Mathematical Models

(VIII.10) A New Approach to Data Analysis

IX Conclusions

APPENDIX B

Overview of Latent Trait Models: Paper Presented at the Fifteenth Annual Meeting of the Behaviormetric Society of Japan, in August, 1987, at Kyushu University, Fukuoka, Japan.

Psychology has its own unique problem of measuring hypothetical constructs, such as ability, attitude, motivation, personality, and so forth. Since they are hypothesized constructs, there is no way to measure them directly, and indirect measurement through individuals' responses to more or less concrete entities, called items, serves an important role. Thus, latent trait models have been developed for measuring hypothetical constructs, especially in the framework of mental test theory, as well as in social attitude measurement.

Let θ be the latent trait, which can be defined either in the unidimensional latent space or in the multidimensional latent space. Let g denote the item, or the smallest entity, responses to which enable us to measure the latent trait indirectly. In mental measurement, for example, the latent trait may be a specified mathematical ability, and an item is a specific question presented in a mathematics test; in social attitude measurement, the latent trait may be the attitude toward a specific social issue, and items are questions incorporated in a questionnaire specifically developed for the purpose. A typical research may start with data collection based upon n items which have been developed for the purpose of measuring a specific hypothetical construct, or latent trait, and the specification of individual differences among human subjects with respect to the specified latent trait may be at the end of the research.

Unlike many other researchers who work in the area of latent trait theory, or item response theory, my interest does not stay solely on the dichotomous response level, on which the item score assumes either 0 or 1 . Years ago, for example, I developed graded response models (Samejima, 1969, 1972), which deal with discrete item responses having more than two item score categories, i.e., $0, 1, \ldots, m_g$, for item g . A general latent trait model for the homogeneous case of the continuous response level, which deals with items having continuous item responses, was also proposed (Samejima, 1973); and later, the normal ogive model was expanded to fit the multidimensional latent space (Samejima, 1974). A direct expansion of the general model for the continuous item response leads us to the situation in which the conditional distribution of the item score, given the latent trait θ , allows to be partly continuous and partly discrete. This general model is for the unidimensional latent space, and includes four different situations, i.e., the closed response situation, the closed/open response situation, the open/closed response situation, and the open response situation.

The operating characteristic of each discrete response, or the operating density characteristic of each continuous response, plays an important role in latent trait theory. The former is the conditional probability of the specific discrete response, given the latent trait. If, for instance, each question in a vocabulary test is scored either correct or incorrect, then the situation belongs to the dichotomous response level, which is a subcategory of the discrete response level. Thus the binary item score, $u_g\ (= 0, 1)$, is assigned to each item response. The item characteristic function $P_g(\theta)$ of a dichotomous item g is defined as

(1) $$P_g(\theta) = \text{Prob.}[u_g = 1 \mid \theta] \quad ,$$

i.e., the conditional probability for the correct answer to item g , given ability θ (cf. Lord and Novick, 1968). Figure 1 illustrates typical monotone increasing item characteristic functions. In this example, these two dichotomous items follow the normal ogive model, whose item characteristic function can be written in the form

(2) $$P_g(\theta) = (2\pi)^{1/2} \int_{-\infty}^{a_g(\theta - b_g)} e^{-u^2/2}\, du$$

with the item discrimination parameter a_g (> 0) and item difficulty parameter b_g . Note that, in the limiting situation where a_g approaches positive infinity, this item characteristic function tends to a step function which is illustrated in Figure 1. This degenerated item characteristic function belongs to the deterministic model known as Guttman Scale, or as biorders.

If each question in a questionnaire is answered by checking one of the four categories, i.e., strongly disagree, disagree, agree, and strongly agree, as is shown in Figure 2, then the situation belongs to the graded response level. This is also a subcategory of the discrete response level (cf. Samejima, 1969, 1972). Thus the graded item score, x_g $(= 0, 1, \ldots, m_g)$, is assigned to each item response, and in this specific example $m_g = 3$. The operating characteristic, $P_{x_g}(\theta)$, of the graded item score x_g is defined as the conditional probability assigned to x_g , given θ . In the normal ogive model, for example, this operating characteristic is given by

$$(3) \qquad P_{x_g}(\theta) = (2\pi)^{-1/2} \int_{a_g(\theta - b_{x_g+1})}^{a_g(\theta - b_{x_g})} e^{-u^2/2}\, du \quad ,$$

where a_g (> 0) is the item discrimation parameter and b_{x_g} is the item response difficulty parameter, which satisfies

$$(4) \qquad -\infty = b_0 < b_1 < \ldots < b_{m_g-1} < b_{m_g} < b_{m_g+1} = \infty \quad .$$

Figure 3 illustrates a set of operating characteristics in the normal ogive model on the graded response level with $a_g = 1.00$, $b_1 = -1.50$, $b_2 = -0.50$, $b_3 = 0.00$, $b_4 = 0.75$ and $b_5 = 1.25$.

An interesting application of the graded response model was made by Roche, Wainer, and Thissen in the area of medical science (Roche, Wainer, and Thissen, 1975). In their research, they developed a skeletal maturity scale, using the knee joint as the biological indicator. There are thirty-four indicators, or items, and through the X-ray films of the subject's knee joint each item was scored by experts into two to five graded categories. Since the subject's skeletal age does not always coincide with his chronological age, the skeletal maturity scale based upon latent trait theory is meaningful.

When the item response is continuous, the operating density characteristic of the continuous item score must be considered. Let z_g be the continuous response. Without loss of generality, we can define z_g as the set of real numbers between zero and unity. In the open response situation, it is assumed that the conditional probability of z_g , given the latent trait θ , is zero for any value of z_g , including the two endpoints, and the conditional distribution of z_g is given as a continuous distribution. As an example, let us consider the response format given as Figure 4. If the subject is asked to respond to a given statement by checking any point in the line segment shown in Figure 4, except for the two endpoints, then it will be reasonable to assume that the conditional probability of any particular point in the line segment is zero, for any fixed θ : thus the open response situation occurs. On the other hand, if the subject is also allowed to check either one of the endpoints, then it will be unreasonable to assume that the conditional probability of each endpoint , i.e., $z_g = 0$ or $z_g = 1$, is zero, for people tend to check either endpoint more often than any other point: thus the closed response situation occurs, and the conditional distribution of z_g , given θ , is continuous for $0 < z_g < 1$, and discrete at $z_g = 0$ and $z_g = 1$.

We notice that in most experimental situations where response latency is used as a measure of "quickness" of information processing, and so on, we are forced to terminate the experiment when the subject's response is too slow. If we consider the response latency as the reversed continuous item score itself by defining the time set as the time limit as zero and "zero second" as the unity, this will be a good example of the closed/open response situation. The conditional probability of z_g , given θ , is zero for any z_g except for $z_g = 0$. Thus the conditional distribution is continuous for $0 < z_g < 1$, and discrete at $z_g = 0$. In a similar manner, the open/closed response situation is defined as the one in which the conditional distribution is continuous for $0 < z_g < 1$ and discrete at $z_g = 1$.

Let $P^*_{z_g}(\theta)$ be the conditional probability with which the subject obtains the item score z_g or greater, given θ . A general mathematical form for $P^*_{z_g}(\theta)$ in the homogeneous case of the continuous response model is given by

$$(5) \qquad P^*_{z_g}(\theta) = \int_{-\infty}^{a_g(\theta - b_{z_g})} \Psi_g(t)\, dt \quad ,$$

with

$$(6) \qquad \begin{cases} \lim_{\theta \to \infty} P^*_{z_g}(\theta) = 0 \\ \lim_{\theta \to -\infty} P^*_{z_g}(\theta) = 1 \quad , \end{cases}$$

where a_g (> 0) is the item discrimination parameter, b_{z_g} is the item response difficulty parameter, and $\Psi_g(\cdot)$ is a specific continuous function, which characterizes the model, and is positive almost everywhere. If, for example, $\Psi_g(t)$ is given by

$$(7) \qquad \Psi_g(t) = (2\pi)^{-1/2} e^{-t^2/2} \quad .$$

then formula (1) will provide us with the normal ogive model, and if it is given by

$$(8) \qquad \Psi_g(t) = De^{-Dt}[1 + e^{-Dt}]^{-2}$$

then the logistic model will be defined.

Some years ago, Birnbaum proposed the logistic model as a good substitute for the normal ogive model on the dichotomous response level (Birnbaum, 1968) because of the similarity of its item characteristic function to the one in the normal ogive model, and also because of the fact that in the logistic model there exists a simple sufficient statistic for the response pattern of binary item scores. It is interesting to note that, as we proceed from the dichotomous response level to the graded response level and, further, to the continuous level, substantial differences between the two models come up to the surface.

The operating density characteristic, $H_{z_g}(\theta)$, has been defined, and it can be written in the form

$$(9) \qquad H_{z_g}(\theta) = a_g \Psi_g\{a_g(\theta - b_{z_g})\}\{\frac{d}{dz_g} b_{z_g}\} \qquad 0 < z_g < 1 \quad .$$

Let $P_{z_g}(\theta)$ be the conditional probability of z_g, given θ. We can write

$$(10) \qquad \int_0^1 H_{z_g}(\theta)\, dz_g = 1 - \{P_0(\theta) + P_1(\theta)\} \leq 1 \quad ,$$

where $P_0(\theta)$ and $P_1(\theta)$ indicate $P_{z_g}(\theta)$ for $z_g = 0$ and $z_g = 1$, respectively. We can also write for the difficulty parameter

$$(11) \qquad \begin{cases} \lim_{z_g \to 0} b_{z_g} = b_0 \geq -\infty \\ \lim_{z_g \to 1} b_{z_g} = b_1 \leq \infty \quad . \end{cases}$$

It is noted that in the open response situation, $P_0(\theta) = P_1(\theta) = 0$, throughout the whole range of θ, and an equality holds in (10) and in each formula of (11). In each of the three situations, however, the left hand side of (10) becomes less than unity, and equality does not hold in both formulae of (11).

Figure 5 illustrates examples of functional relationships between the continuous item score z_g and the difficulty parameter b_{z_g} in the closed response situation. The two functional formulae used in those examples are given in the caption of the figure. In the closed/open response situation, those curves approach positive infinity as z_g tends to unity; in the open/closed response situation, they approach negative infinity as z_g tends to zero; and in the open response situation, both of these asymptotic characteristics must be true.

Figure 6 illustrates the operating density characteristic, $H_{z_g}(\theta)$, in the normal ogive model and in the logistic model for the five selected values of z_g in the closed response situation, using the linear difficulty parameter function illustrated in Figure 5. As you can see in this figure, for each and every item score θ for $0 < z_g < 1$, the operating density characteristic has a unique local maximum, and the configurations of those curves in the two separate models are similar.

So far I have attempted to present a rough outline of latent trait theory, selecting several representative situations and models, among others. An important basic characteristic of this very comprehensive theory is that it is a probabilistic model, not a deterministic model. Estimation of the operating characteristics, or of the operating density characteristics, of the item response is, therefore, one of the most important objectives in the methodology of latent traity theory. There have been many methods and computer programs developed by researchers in this area. Those methods can be categorised into two categories, i.e. 1) the parametric method, and 2) the nonparametric method. In the former, we assume a certain mathematical model for the operating characteristic, and the estimation is reduced to that of the item parameters. In the latter, we attempt to approach the operating characteristic directly, avoiding assumptions as much as possible. I have developed several methods and approaches in the past eight years, which are categorized into the nonparametric method, in the multiyear research projects supported by the Office of Naval Research (cf. Samejima, 1981). These methods and approaches, which are listed below, are basically for the discrete item responses in the unidimensional latent space, although they can be applied for the continuous item responses.

(1) Approaches
- (i) Bivariate P.D.F. Approach
- (ii) Histogram Ratio Approach
- (iii) Curve Fitting Approach
- (iv) Conditional P.D.F. Approach
  - (a) Simple Sum Procedure
  - (b) Weighted Sum Procedure
  - (c) Proportioned Sum Procedure

(2) Methods
- (i) Pearson System Method
- (ii) Two-Parameter Beta Method
- (iii) Normal Approach Method

Out of those combinations of an approach and a method, Simple Sum Procedure of the Conditional P.D.F. Approach which is combined with the Normal Approach Method has been used most frequently for analysing empirical data. We assume that we have a set of n items whose operating characteristics are known, which is called Old Test following the terminology in mental measurement. For convenience, let us assume that each item g of the Old Test belongs to the graded response level. Let $I_{x_g}(\theta)$ denote the item response information function, which is defined by

$$I_{x_g}(\theta) = -\frac{\partial^2}{\partial\theta^2}\log P_{x_g}(\theta) \quad . \tag{12}$$

The item information function, $I_g(\theta)$, is given as the conditional expectation of the item response information function, given θ, i.e.,

$$I_g(\theta) = \sum_{x_g=0}^{m_g} I_{x_g}(\theta)P_{x_g}(\theta) \quad . \tag{13}$$

The response pattern, V , of those n items of the Old Test can be written as

(14) $$V = (x_1, x_2, \ldots, x_g, \ldots, x_n)' \ .$$

By virtue of the conditional independence of the item score distributions (Lord and Novick, 1968), the operating characteristic, $P_V(\theta)$, of the response pattern V is given by

(15) $$P_V(\theta) = \prod_{x_g \in V} P_{x_g}(\theta) \ .$$

This operating characteristic, $P_V(\theta)$, is also the likelihood function, $L_V(\theta)$, for estimating the individual parameter θ_s for individual s from his response pattern, when the item parameters are known. We can write for the response pattern information function, $I_V(\theta)$,

(16) $$I_V(\theta) = -\frac{\partial^2}{\partial\theta^2} \log P_V(\theta) = \sum_{x_g \in V} I_{x_g}(\theta) \ ,$$

and the test information function $I(\theta)$ is defined as the conditional expectation of the response pattern information function, given the latent trait θ . Thus we have

(17) $$I(\theta) = \sum_V I_V(\theta) P_V(\theta) = \sum_{g=1}^{n} I_g(\theta) \ .$$

Figure 5 illustrates the square root of the test information functions of the Level 11 Vocabulary Subtest of the Iowa Tests of Basic Skills. In order to make the amount of information equal for the interval of latent trait of interest, θ is transformed to τ by

(18) $$\tau = C_1^{-1} \int_{-\infty}^{\theta} [I(t)]^{1/2}\, dt + C_0 \ ,$$

where C_0 is an arbitrary constant for adjusting the origin of τ , and C_1 is another arbitrary constant which equals the square root of the constant test information function, $I^*(\tau)$, of τ .

Using the symptotic property of the maxumum likelihood estimate, which distributes normally with the true parameter θ and the inverse of the square root of the test information as its two parameters, as an approximation to the conditional distribution of $\hat{\tau}$, given τ , we obtain for the two conditional moments

(19) $$E(\tau \mid \hat{\tau}) = \hat{\tau} + C_1^{-2} \frac{d}{d\hat{\tau}} \log g^*(\hat{\tau})$$

and

(20) $$\mathrm{Var.}(\tau \mid \hat{\tau}) = C_1^{-2}[1 + C_1^{-2} \frac{d^2}{d\hat{\tau}^2} \log g^*(\hat{\tau})] \ .$$

In the Simple Sum Procedure of the Conditional P.D.F. Approach, the estimate of the operating characteristic, $P_{k_h}(\theta)$, of the discrete item response k_h to the "unknown" item h is given by

(21) $$\hat{P}_{k_h}(\theta) = \hat{P}^*_{k_h}[\tau(\theta)] = \sum_{s \in k_h} \hat{\phi}(\tau \mid \hat{\tau}_s)[\sum_{s=1}^{N} \hat{\phi}(\tau \mid \hat{\tau}_s)]^{-1} \ ,$$

where N is the number of individuals in our sample and $\hat{\phi}(\tau \mid \hat{\tau}_s)$ is the estimated conditional density of τ, given the maximum likelihood estimate $\hat{\tau}_s$ of the individual s. In the Normal Approach Method, this conditional density is approximated by the normal density using the two estimated parameters derived from (19) and (20) by setting $\hat{\tau} = \hat{\tau}_s$.

Those methods and approaches for estimating the operating characteristics of discrete item responses can effectively be applied for the on-line item calibration in computerized adaptive testing. The idea of adaptive testing is to increase the efficiency of ability estimation, by presenting an optimal subset of small number of items selected from a large item pool to an individual subject. With the rapid progress of computer technology, computerized adaptive testing has become more and more popular in the past decade. In adaptive testing, it is necessary to add new items to the item pool as we continue using it for years. We can use these nonparametric methods and approaches for the on-line calibration of new items, if we use and appropriate constant amount of test information to all the individuals as the criterion for terminating the presentation of items from the "old" item pool. Note that in this example, Old Test does not consist of a single set of items, and yet the test information of the Old Test is kept constant, so that we do not need to transform θ to τ, and, as the result, the estimation procedure becomes much simplified.

As we can see from the above examples of skeletal maturity, reaction time, on-line calibration in computerized adaptive testing, and so forth, latent trait theory has a very broad area of conceivable applications. As a researcher who has been working on latent trait theory and its methodologies for many years, I wish to see many more applications of the theory in the area of natural sciences, as well as social sciences, in the future.

References

[1] Lord, F.M. and M.R. Novick. *Statistical theories of mental test scores*. Addison-Wesley, 1968, Chapter 16.

[2] Roche, A., Wainer, H., and Thissen, D. *Skeletal Maturity: The Knee Joint as a Biological Indicator*. New York: Plenum Medical Book Company, 1975.

[3] Samejima, F. Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17, 1969.

[4] Samejima, F. A general model for free-response data. *Psychometrika Monograph*, No. 18, 1972.

[5] Samejima, F. Homogeneous case of the continuous response level. *Psychometrika*, 1973, 38, 203-219.

[6] Samejima, F. Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika*, 1974, 39, 111-121.

[7] Samejima, F. A new family of models for the multiple-choice item. *Office of Naval Research*, Research Report 79-4, 1979.

[8] Samejima, F. Final Report: Efficient methods of estimating the operating charactristics of item response categories and challenge to new model for the multiple-choice item. *Final Report of N00014-77-C-0360, Office of Naval Research*, 1981.

[9] Samejima, F. Plausibility Functions of Iowa Vacabulary Test Items Estimated by the Simple sum Procedure of the Conditional P.D.F. Approach. *Office of Naval Research*, Research Report 84-1, 1984.

APPENDIX C

Ten Items of the Rorschach Test and Their Scorings for the Purpose of Measuring the Intellectural Capability Using Appropriate Latent Trait Models

(1) Failure to Justify in the Abscence of Imaginal Aspects (Qualities) (f) (0 - 1)

1 - [(total number of failures in justifying non-imaginal responses) / (total number of non-imaginal responses)]

NOTE: When the total number of non-imaginal responses is 4 or less, we exclude this item for the subject in question.

Example 1.1: Imaginal response.
"It looks like a river way down at the bottom of a valley."

Example 1.2: Non-imaginal response.
"It looks like a bat."

Example 1.3: Failure in justifying the above non-imaginal response.
Example 1.2 plus: "It just looks like it."

Example 1.4: Success in justifying the above non-imaginal response.
Example 1.2 plus: "Because of the way it's shaped."

Model: Discrete/Continuous Model, Closed Response Situation.

(2) Animal Ratio (ANRA) (0 - 1)

1 - [(total number of animal responses without humans) / (total number of responses)]

NOTE: One response can contain 2 or more content scorings. e.g.; "A man walking a dog." (2 content scorings: H and A) Because this includes human, it is not considered as an "animal response", and is counted as 0 in the numerator of the above ratio.

When the total number of responses is 11 or less, then we exclude this item for the subject in question.

This rule applies for any item using the total number of responses in the denominator. In fact, for such a subject

Rorschach diagnosis is practically impossible.

Approximately 50% of adults' responses are animal responses.

Example 2.1: Animal response without humans.
"This looks like a bat."
"A dog standing on a river bank."

Model: Discrete/Continuous Model, Closed Response Situation.

(3a) Response Complexity (RESCOM) (0 - 1)

(1/8)*[(number of justification types + number of imaginal aspects) for each response summed over all responses] / (total number of responses) if B = 8 or less
1 if B is more than 8

where B denotes the number of elements in a blend for each response.

NOTE: 1 response can contain up to 13 justification types and imaginal aspects. Since it is rare to have more than 8 elements in a "blend", and in actual diagnosis clinicians do not usually make differences between 8 and 13, they are included in a single item score, 1 .

Hereafter, we use "elements in a blend" for (number of justification types + number of imaginal aspects). When abbreviated, justifications (alphabetized) are followed by imaginal aspects (alphabetized).
e.g.; C.C'.F.HE.HM.V (color, achromatic color, form, human emotion, human movement, vista)

Example 3a.1: F (form) (1 element)
"It looks like a landscape, because of the shape."

Example 3a.2: C'.F (2 elements)
(The above plus:) "And the white could be snow."

Example 3a.3: C.C'.F (3 elements)
(The above plus:) "And the red looks like a campfire."

Example 3a.4: C.C'.F.V (4 elements)
(The above plus:) "And in the distance a man."

Example 3a.5: C.C'.F.HE.V (5 elements)
(The above plus:) "He looks happy."

Example 3a.6: C.C'.F.HE.HM.V (6 elements)

(The above plus:) "He is cooking."

Example 3a.7: Ca (arbitrary color) (1 element)
"It's a map, because maps are always colored."

Example 3a.8: Cp (projected color) (1 element)
(To a non-colored blot:) "It's a blue bird because it's blue."

Example 3a.9: Sh (shading) (1 element)
"It's like fog that you can almost see through."

Example 3a.10: T (texture) (1 element)
"It looks like it would feel soft."

Example 3a.11: AM (animal movement) (1 element)
"It looks like a dog barking."

Example 3a.12: OM (object movement) (1 element)
"It looks like a volcano exploding."

Example 3a.13: Ci (inappropriate color) (1 element)
"It looks like a green sheep."

Model: Discrete/Continuous Model, Open/Closed Response Situation.

(3b) Maximum Response Complexity (MRC) (0 - 1)

max{B} / 8	if max{B} = 8 or less
1	if max{B} is more than 8

Model: Discrete/Continuous Model, Open/Close Response Situation.

(3c) Proportion of Complex Blends (PCB) (0 - 1)

(number of responses with 4 or more elements in a blend) / (number of total responses)

NOTE: 3 or less elements in a blend cannot be considered as many, so we take 4 or more to indicate "many" for each response.

Model: Discrete/Continuous Model, Closed Response Situation.

(4) Corrected Total Number of Responses (CRES) (0 - 1)

(total number of responses) / 50	if R = 50 or less
1	if R is more than 50

Model: Discrete/Continuous Model, Open/Closed Response Situation.
If the total number of responses are, say, 50 or 99, the difference is not counted in the diagnosis situation.

(5) Mean Human Articulation (MEHA) (0 - 1)

Take the mean of the distribution of HA1, HA2, HA3 and HA4, giving 0, 1/3, 2/3 and 1 for the separate scores. There is some doubt that HA4 is not exactly ordered, but it has been decided to take the present policy at this stage, and we will come back to this point later in the analysis.

Example 5.1: HA1. "It looks like a person."

Example 5.2: HA2. "It looks like a big person."

Example 5.3: HA3. "It looks like a big man."

Example 5.4: HA4. "It looks like a policeman."

Model: Discrete/Continuous Model, Closed Response Situation.

(6) Mean Cognitive Complexity (MECOG) (0 - 1)

We have discussed up to the point to decide that the categories should be ordered as: 1) simple, 2) diffuse, 3) articulated + arbitrary and 4) integrated.

Take the mean of the distribution of the above 4 categories, giving 0, 1/3, 2/3 and 1 for the separate scores for the above 4 categories. There is some doubt about this measure, for one of the clinicians says that she does not take the frequencies of 1) and 2) into consideration when she diagnoses. We will come back to this question of the adequacy of this item later in the analysis.

Example 6.1: Simple cognition.
"It looks like a bat."

Example 6.2: Diffuse cognition.
"It looks like a cloud."

Example 6.3: Articulated cognition.
"It looks like a chair with a back and four legs."

Example 6.4: Arbitrary cognition.
"It looks like a spider wearing a hair ribbon."

Example 6.5: Integrated cognition.
"It looks like two people dancing together."

Model: Discrete/Continuous, Closed Response Situation.

(7) Proportion of Pure Form Justifications That Are Socially Appropriate (F+%) (0 - 1)

(number of F+ justifications) / (total number of F justifications)

NOTE: Pure form justifications (F) includes both F+ and F- , i.e., pure form justifications that are socially inappropriate. F is the most common justification, and approximately 50% of the adult subjects' justifications belong to this category.

Categorization of F responses into F+ and F- categories is made following the list in Beck, S. J., et al, 1961.

If the total number of F justifications is 4 or less, the item is excluded for that subject. Typically, for one subject there are 10 to 12 F justifications.

Example 7.1: F+ justification.
(To the whole area of the inkblot of card 1 which really looks like a bird:)
"Bird, because it is shaped like a bird."

Example 7.2: F- justification.
(To the same area indicated above:)
"Worm, because it is shaped like a worm."

Model: Discrete/Continuous Model, Closed Response Situation.

(8) Proportion of Whole Responses (W%) (0 - 1)

(number of responses using the whole inkblot) / (total number of responses)

NOTE: Approximately 20% of responses are whole responses in the adult population.

There is some concern that the relationship of this item with the intellectual capability may not be monotonic. We will come back to this question later in the analysis.

Model: Discrete/Continuous, Closed Response Situation.

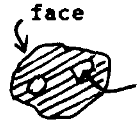
(9) Integrated White Space (IWS) (0,1,2,...)

(number of responses with integrated white space)

NOTE: Usually, black and/or colored parts of the picture (inkblot) are used as figure against a white ground. But sometimes white spaces are integrated in the figure.

6 or more such responses are rare, and usually in diagnosis they are considered in a single category, i.e., "many".

Example 9.1:



Model: Graded Response Model.

(10) Range of Content (CONTRA) (0,1,2,...,14,many)

(number of different content categories scored)

NOTE: This starts from 1 and goes up to 37. 15 or more are rare. They may be categorized as "many".

These categories include: 1) animal, 2) special case of animal (e.g.; dragon), 3) abstraction (e.g.; conflict), 4) animal detail (e.g.; dog's paw), 5) special case of animal detail (e.g.; phoenix' wing), 6) animal face , 7) special case of animal face (e.g.; unicorn's face), 8) anatomy, bony, 9) anatomy, soft (lung), 10) anatomy, sex, clothing (e.g.; bow tie), 16) cloud, smoke, vapor, 17) death (e.g.; tombstone), 18) emblems (e.g.; flag), 19) food, 20) fire (including explosion), 21) geography (e.g.; Italy), 22) human, 23) special case of human (e.g.; demon), 24) human detail (excluding human head or face, e.g.; finger), 25) special case of human detail (e.g.; mermaid's tail), 26) human face, 27) special case of human face (ghost's face), 28) household (e.g.; drawer cabinet), 29) implements (e.g.; hammer), 30) landscape, 31) music (e.g.; violin), 32) religion (e.g.; cross), 33) schemata (e.g.; map), 34) scientific implements (e.g., microscope), 35) toy (e.g.; bicycle), 36) travel (e.g., airplane), 37) weapons (e.g.; missile). (cf. Burstein and Loucks, 1988)

Model: Graded Response Model.

REMARKS: (3a), (3b) and (3c) are included because they are all used in actual clinical diagnosis for intellectual ability. In the process of analysis, however, they may be combined into one item.

Motivational articulation was also considered as an item, but excluded later, since we discovered that there is no systematic tendency to believe that MA1, MA2, MA3 and MA4 are ordered to reflect intellectual capacity.

(1) Beck, S. J., Beck, A. G., Levitt, E. E. and Molish, H. B. Rorschach Test, Vol. 1: Basic Process. New York: Grune and Stratton, 1961.

(2) Burstein, A. G. and Loucks, S. Rorschach's Test: Scoring and Interpretation. New York: Hemisphere. (To be published in August, 1988.)

RORSCHAC

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

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

Dr. James Algina
University of Florida
Gainesville, FL 32605

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

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

Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

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

Dr. Arthur S. Blaiwes
Code N711
Naval Training Systems Center
Orlando, FL 32813

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

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

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

Dr. Robert Breaux
Code N-095R
Naval Training Systems Center
Orlando, FL 32813

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

Dr. Lyle D. Broemeling
ONR Code 1111SP
800 North Quincy Street
Arlington, VA 22217

Mr. James W. Carey
Commandant (G-PTE)
U.S. Coast Guard
2100 Second Street, S.W.
Washington, DC 20593

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

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

Dr. Robert Carroll
OP 01B7
Washington, DC 20370

Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

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

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

Dr. Hans Crombag
University of Leyden
Education Research Center
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

Dr. Timothy Davey
Educational Testing Service
Princeton, NJ 08541

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

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

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

Dr. Hei Ki Dong
Bell Communications Research
6 Corporate Place
PYA-1k226
Piscataway, NJ 08854

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

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

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

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

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

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

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

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

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

Dr. Pat Federico
Code 511
NPRDC
San Diego, CA 92152-6800

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

Dr. Richard L. Ferguson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52240

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

Dr. Myron Fischl
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

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

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

Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

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

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

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

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

Dipl. Pad. Michael W. Habon
Universitat Dusseldorf
Erziehungswissenschaftliches
Universitatsstr. 1
D-4000 Dusseldorf 1
WEST GERMANY

Dr. Ronald K. Hambleton
Prof. of Education & Psychology
University of Massachusetts
at Amherst
Hills House
Amherst, MA 01003

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

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

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

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

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

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 90010

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

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

Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

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

Dr. Robert Jannarone
Department of Psychology
University of South Carolina
Columbia, SC 29208

Dr. Dennis E. Jennings
Department of Statistics
University of Illinois
1409 West Green Street
Urbana, IL 61801

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

Dr. Milton S. Katz
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

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

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

Dr. William Koch
University of Texas-Austin
Measurement and Evaluation
Center
Austin, TX 78703

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

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

Dr. Daryll Lang
Navy Personnel R&D Center
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

Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

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

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

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

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

Dr. William L. Maloy
Chief of Naval Education
and Training
Naval Air Station
Pensacola, FL 32508

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

Dr. Clessen Martin
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

Dr. James McBride
Psychological Corporation
c/o Harcourt, Brace,
Javanovich Inc.
1250 West 6th Street
San Diego, CA 92101

Dr. Clarence McCormick
HQ, MEPCOM
MEPCT-P
2500 Green Bay Road
North Chicago, IL 60064

Dr. Robert McKinley
Educational Testing Service
20-P
Princeton, NJ 08541

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

Dr. Barbara Means
Human Resources
Research Organization
1100 South Washington
Alexandria, VA 22314

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

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

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

Headquarters, Marine Corps
Code MPI-20
Washington, DC 20380

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

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

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

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

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

Fleet Support Office,
NPRDC (Code 301)
San Diego, CA 92152-6800

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

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

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

Dr. James Olson
WICAT, Inc.
1875 South State Street
Orem, UT 84057

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

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

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

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

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

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

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

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

Administrative Sciences Department,
Naval Postgraduate School
Monterey, CA 93940

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

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

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

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

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

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

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
3108 AustinPeay Bldg.
Knoxville, TN 37916-0900

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

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

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

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

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

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

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

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

Dr. Richard E. Snow
Department of Psychology
Stanford University
Stanford, CA 94306

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

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

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

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

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
San Diego, CA 92152-6800

Dr. John Tangney
AFOSR/NL
Bolling AFB, DC 20332

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
Urbana, IL 61801

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

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

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

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

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

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

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 310
St. Paul, MN 55114

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

Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08541

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

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

Dr. Brian Waters
Program Manager
Manpower Analysis Program
HumRRO
1100 S. Washington St.
Alexandria, VA 22314

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

Dr. Ronald A. Weitzman
NPS, Code 54Wz
Monterey, CA 92152-6800

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90007

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

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

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

Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152-6800

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
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
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
Memory & Cognitive
Processes
National Science Foundation
Washington, DC 20550

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

NON GOVT

1 Dr. Robert Guion
Department of Psychology
Bowling Green State University
Bowling Green, OH 43403

1 Dr. P. Mengal
Faculte de Psychologie
et des Sciences de I'Education
Universite de Geneva
3 fl. de I'Universite
1201 Geneva SWITZERLAND

1 Dr. Wim J. van der Linden
Vakgroep Onderwisjskunde
Postbus 217
7500 EA Endschede
THE NETHERLANDS

1 Dr. Lutz Hornke
University Duesseldorf
Erz. Wiss.
D-4000 Duesseldorf
WEST GERMANY

1 Dr. Wolfgang Buchtala
8346 Simbach Inn
Postfach 1306
Industriestasse 1
WEST GERMANY

1 Dr. Albert Beaton
Educational Testing Service
Princeton, NJ 08542

1 Dr. Ivo W. Molenaar
F.S.W.-R.U.G.
Oude Boteringerstraat 23
9712 GC Groningen
THE NETHERLANDS

1 Dr. Chang-I Bonnie Chen
Graduate School of Psychology
National Chenglhi University
Taipei, Taiwan
R.O.C.

NON GOVT

1 Dr. Sukeyori Shiba
Faculty of Education
University of Tokyo
Hongo, Bumkyoku
Tokyo, JAPAN 113

1 Dr. Takahiro Sato
Nippon Electric Co., Ltd
C & C Information Technology
Research Laboratories
1-1 Miyazaki 4-Chome
Miyamae-ku, Kawasaki
Kanagawaken 213, JAPAN

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Ave.
Woodland Hills, CA 91634

1 Mr. Kenji Goto
Ina Nyuhaim #208
7-2-12 Honcho, Tanashi-shi
Tokyo 188, JAPAN

1 Dr. G. Gage Kingsbury
Portland Public Schools
Evaluation Department
501 North Dixon Street
Portland, Oregon 97227

1 Mr. Tadashi Shibayama
Ono-kohpo #201
901 Hiregasaki
Nagareyama-shi
Chiba-ken, 270-01
JAPAN

NAVY

1 Mr. Thomas Bryant
Office of Naval Research
206 O'Keefe Building
Atlanta, GA 30332

ARMY

1 Dr. Randall M. Chambers
U. S. Army Research Institute
for the Behavioral & Social Sciences
Fort Sill Field Unit
P. O. Box 3066
Fort Sill, OK 73503

END
DATED
FILM
8-88
DTIC