

COGNITIVE DIAGNOSIS COMBINED WITH LATENT TRAIT THEORY AND MULTIPHASE RESPONSES

FUMIKO SAMEJIMA

UNIVERSITY OF TENNESSEE
KNOXVILLE, TENN. 37996-0900

NOVEMBER, 1994

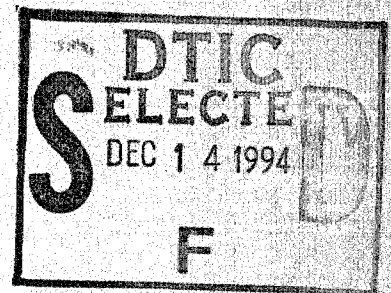
Prepared under the contract number N00014-90-J-1456,
4421-549 with the
Cognitive Science Research Program
Cognitive and Neural Science Division
Office of Naval Research

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Government.

R01-1069-25-001-95

19941209 098

DTIC QUALITY INSPECTED 1



REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 25 Nov. 94	3. REPORT TYPE AND DATES COVERED 1 Jan. 90 - 31 Aug. 94	
4. TITLE AND SUBTITLE Cognitive Diagnosis Combined with Latent Trait Theory and Multiphase Responses			5. FUNDING NUMBERS N00014-90-J-1456 PE 61153N PR RR-042-04 TA 042-04-01 WU 4421-549	
6. AUTHOR(S) Fumiko Samejima, Ph.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Fumiko Samejima, Ph.D. Department of Psychology University of Tennessee 310B Austin Peay Knoxville, TN 37996-0900			8. PERFORMING ORGANIZATION REPORT NUMBER R01-1069-25-001-95	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Cognitive Science 342 CS Office of Naval Research 800 North Quincy Street Arlington, VA 22217			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release Distribution unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This is a summary of the research conducted in 1900-94 under the title, "Cognitive Diagnosis Combined with Latent Trait Theory and Multiphase Responses."				
14. SUBJECT TERMS Cognitive Diagnosis, Graded Response Model, Latent Trait Models, Item Response Theory, Nonparametric Estimation			15. NUMBER OF PAGES 46	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT Unlimited	

PREFACE

Four years and eleven months have passed since this research period started on January 1, 1990. During this period, effort was focused on:

1. further development of theory and methodologies in latent trait models, including the proposal of a new model, called acceleration model,
2. development of methodologies for cognitive diagnosis assessment applying and expanding latent trait models, which were eventually integrated into the method, called competency space approach,
3. theoretical integration of the nonparametric approaches and methods developed in the past years, and
4. writing research outcomes in refereed journal papers and book chapters, and presenting them at international and domestic conferences.

During the research period there were many people on the University of Tennessee Knoxville campus, including the Acting Director, Mr. Bruce H. Delaney, of the Computing Center, who helped me in conducting research; their helps are highly appreciated. Also I would like to express my gratitude to people of the Office of Naval Research, especially the scientific officers Dr. Charles E. Davis and Dr. Susan E. Chipman and the ONR representatives in Atlanta, including Mr. Thomas Bryant and Ms. C. C. Everley.

Special thanks are due to my graduate assistant, Mr. Christopher Coleman, for helping me in preparing this final research report.

Accession For	
NTIS CRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Date Rec'd /	
Source	
Agency / Office	
Project Title / Special	
A-1	

November 25
Author

TABLE OF CONTENTS

I. Introduction	1
[I.1] Book Chapters	1
[I.2] Refereed Journal Papers	1
[I.3] Proceedings Article	2
[I.4] Papers in Preparation for Refereed Journals	2
[I.5] Invited Paper Presentations at International Conferences	2
[I.6] Invited Paper Presentations at Domestic Conferences	3
[I.7] Contributed Paper Presentations at Domestic Conferences	3
[I.8] Seminars	4
[I.9] Awards	4
II. Competency Space Approach to Cognitive Diagnosis	5
[II.1] Method	5
[II.2] Domain Knowledge "Plus"	5
[II.3] Preliminary Study: Two Quizzes	6
[II.4] Different Types of Test Items	6
[II.4.1] PAPER-AND-PENCIL TESTS VS. COMPUTERIZED TESTS	6
[II.4.2] RECOGNITION VS. CONSTRUCTION	6
[II.4.3] PROBLEM COMPLEXITY	7
[II.4.4] CONVENTIONAL TESTS VS. ADAPTIVE TESTS	7
[II.5] Computerized Tests with <Given>, <Hint> and <Terminal>	8
[II.6] Diffused Attributes and Concentrated Attributes	9
[II.7] Alternative Method: Multi-Stage Latent Trait Approach	11
[II.8] Grades of Attainment	13
[II.9] Decomposition of the Competency Space	14
III. Efficient Nonparametric Approaches for Estimating the Operating Characteristics of Discrete Item Responses	16
[III.1] Rationale	17
[III.2] Bivariate P.D.F. Approach	19
[III.3] Conditional P.D.F. Approach	22
[III.3.1] SIMPLE SUM APPROACH	22
[III.3.2] DIFFERENTIAL WEIGHT APPROACH	25

IV. Acceleration Model	29
[IV.1] Processing Functions	29
[IV.2] Criteria for Evaluating Mathematical Models	30
[IV.3] General Acceleration Model and a Specific Model in Which the Logistic Function is Used	31
V. Further Research and Integration of Research Findings	34
[V.1] Further Research	34
[V.1.1] MLE BIAS FUNCTION	34
[V.1.2] CRITICAL OBSERVATIONS OF THE TEST INFORMATION FUNCTION AS A MEASURE OF LOCAL ACCURACY	34
[V.1.3] PLAUSIBILITY FUNCTIONS OF DISTRACTORS	34
[V.1.4] ESTIMATION OF RELIABILITY COEFFICIENTS USING THE TEST INFORMATION FUNCTION AND ITS MODIFICATIONS	34
[V.2] Integration of Research Findings	35
[V.2.1] ROLES OF FISHER TYPE INFORMATION IN LATENT TRAIT MODELS	35
[V.2.2] HUMAN PSYCHOLOGICAL BEHAVIOR	35
[V.2.3] GRADED RESPONSE MODEL	35
VI. Discussion	37

I. Introduction

Roughly speaking, in the first half of the research period, the principal investigator's effort was mainly focused on developing theories and methodologies for cognitive diagnosis based on latent trait models, interacting with Drs. Susan Goldman, Gautum Biswas and other researchers of Vanderbilt University, Nashville, Tennessee, who worked on trouble shootings in complementary metal oxide semiconductor (CMOS) design tasks; in the second half of the research period, emphasis was put upon writing and publishing book chapters and papers for refereed journals on various research topics which were investigated in the years of the Office of Naval Research fundings that started in 1977, *adding further research*, while continuing developing theories and methodologies for cognitive assessment. A latent trait model, called *acceleration model*, was proposed in the second half, which belongs to the heterogeneous case of the graded response model (Samejima, 1972), and will be useful in cognitive assessment as well as in more traditional areas that latent trait models have been applied for, such as mental testing.

During the research period, 1990-94, 3 book chapters were written, one of which was published, and the other two are in press; 7 papers were published or accepted for publication in refereed journals; 1 paper was published in international conference proceedings; and two papers were prepared for submission to refereed journals. The titles of these papers are as follows:

[I.1] Book Chapters

- [1] Roles of Fisher Type Information in Latent Trait Models, in H. Bozdogan (Ed.), *Proceedings on the First US/Japan Conference on: The Frontiers of Statistical Modeling: An Informational Approach* (3 volumes), Netherlands: Kluwer Academic Publishers. 1994.
- [2] A Cognitive Diagnosis Method Using Latent Trait Models: Competency Space Approach and its Relationship with Dibello and Stout's Unified Cognitive/Psychometric Diagnosis Model, in P. D. Nichols, S. E. Chipman & R. L. Brennan (Eds.), *Cognitive Diagnostic Assessment*, Hillsdale, New Jersey: Lawrence Erlbaum. (in press).
- [3] Graded Response Model, in W. J. van der Linden & R. Hambleton (Eds.), *Handbook of Modern Item Response Theory*, New York: Springer-Verlag. (in press).

[I.2] Refereed Journal Papers

- [4] An approximation for the bias function of the maximum likelihood estimate of a latent variable for the general case where the item responses are discrete, *Psychometrika*, 58, 119-138, 1993.

- [5] The bias function of the maximum likelihood estimate of ability for the dichotomous response level, *Psychometrika*, 58, 195-209, 1993.
- [6] Some critical observations of the test information function as a measure of local accuracy in ability estimation, *Psychometrika*, 59, 307-329, 1994.
- [7] Acceleration model in the heterogeneous case of the general graded response model, *Psychometrika*, (in press).
- [8] Nonparametric estimation of the plausibility functions of the distractors of vocabulary test items, *Applied Psychological Measurement*, (in press).
- [9] Estimation of reliability coefficients using the test information function and its two modifications, *Applied Psychological Measurement*, (in press).
- [10] Efficient nonparametric approaches for estimating the operating characteristics of discrete item responses. (accepted by *Psychometrika*).

[I.3] Proceedings Article

- [11] Human psychological behavior: viewed from latent trait models. In *Proceedings of the IEEE International Workshop on Neuro Fuzzy Control*, March 22-23, 1993, Muroran Institute of Technology, Muroran, Japan.

[I.4] Papers in Preparation for Refereed Journals

- [12] Proposal of two modification formulas of the test information function.
- [13] A latent trait model for the continuous item response whose distribution is partly discrete.

During this period, the principal investigator presented 2 invited papers at international conferences, 3 invited papers at domestic conferences, and 12 contributed papers at domestic conferences, as follows:

[I.5] Invited Paper Presentations at International Conferences

- [1] *Roles of Fisher type information in latent trait models*. US/Japan Conference on the Frontiers of Statistical Modeling: An Information Approach, Knoxville, Tennessee, May 1992.
- [2] *Human psychological behavior: viewed from latent trait models*. The IEEE International Workshop on Neuro-Fuzzy Control, Muroran, Hokkaido, Japan, March 1993.

[I.6] Invited Paper Presentations at Domestic Conferences

- [3] *Partial credit model and extensions*, (discussant paper). The Annual Meeting of the American Educational Research Association, Division D. Atlanta, Georgia; April, 1993.
- [4] *A cognitive diagnosis method using latent trait models: competency space approach and its relationship with DiBello and Stout's unified cognitive/psychometric diagnosis model*. ACT/ONR Conference on Alternative Diagnosis Assessment. Iowa City, Iowa; May, 1993.
- [5] *Assessing dimensionality in item response theory models*, (discussant paper). The Annual Meeting of the National Council on Measurement in Education. New Orleans, Louisiana; April, 1994.

[I.7] Contributed Paper Presentations at Domestic Conferences

- [6] *Differential weight procedure, a nonparametric approach for estimating the operating characteristics of discrete item responses*. The Annual Meeting of the American Educational Research Association. Chicago, Illinois; April, 1991.
- [7] *Some considerations for the refinement of differential weight procedure of the conditional p.d.f. approach*. ONR Conference on Model-Based Measurement. Princeton, New Jersey; May, 1991.
- [8] *An efficient nonparametric method for estimating the operating characteristics of discrete responses*. The Annual Meeting of the Psychometric Society. New Brunswick, New Jersey; June, 1991.
- [9] *Usefulness of latent trait approach in cognitive diagnosis*. ONR Conference on Cognitive Diagnosis. Pittsburg, Pennsylvania; October, 1991.
- [10] *Prospects of applications of nonparametric methods of estimating the operating characteristics in educational measurement* (round table session). The Annual Meeting of the National Council on Measurement in Education. San Francisco, California; April, 1992.
- [11] *Comparisons of the estimated operating characteristics obtained by the simple sum and the differential weight procedures both in conventional and adaptive testings*. The Annual Meeting of the American Educational Research Association. San Francisco, California; April, 1992.
- [12] *A Design of Cognitive Diagnosis*. ONR Conference on Cognitive Diagnosis/Model-Based Measurement. Champaign, Illinois; June, 1992.

- [13] *Two modification formulae of the test information function based upon the MLE bias function.* The Annual Meeting of the Psychometric Society. Columbus, Ohio; July, 1992.
- [14] *Reliability coefficient and standard error of measurement viewed from latent trait models and their effective use.* The Annual Meeting of the American Educational Research Association. Atlanta, Georgia; April, 1993.
- [15] *Comparison of the predicted reliability coefficients using the test information function and its two modification formulae.* The Annual Meeting of the Psychometric Society. Berkeley, California; July, 1993.
- [16] *Cognitive diagnosis using latent trait models,* in the symposium, *Approaches to Cognitive Modeling.* The Annual Meeting of the National Council on Measurement in Education. New Orleans, Louisiana; April, 1994.
- [17] *Acceleration model: a family of graded response or partial credit models in the heterogeneous case.* The Annual Meeting of the Psychometric Society. Champaign, Illinois; June, 1994.

During this period, the principal investigator gave two seminars with herself as a sole speaker, as follows:

[I.8] Seminars

- [1] *Acceleration model and its use in the competency space approach for cognitive diagnosis.* Educational Testing Service, Princeton, New Jersey; March, 1994.
- [2] *Item response theory.* Tokyo Institute of Technology, Tokyo, Japan; August, 1994.

During this period, the principal investigator was awarded by the following.

[I.9] Awards

- [1] Appointment on the *Board of Trustee of the Psychometric Society.* (1989-)1990.
- [2] *Outstanding Technical Contribution Award from the National Council on Measurement in Education,* at the Annual Meeting of the National Council on Measurement in Education, Chicago, Illinois, April, 1991.

References

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

II. Competency Space Approach to Cognitive Diagnosis

A comprehensive methodology for cognitive diagnosis has been developed (Samejima, 1991, 1992, 1994b, 1994e) by systematically using intensive observations of subjects' behavior and advanced theories and methodologies originally developed in psychometrics, taking advantage of advanced computer technologies. For convenience, in this section, *circuit design* is used as the task the examinee works on.

[II.1] Method

The following three steps are taken in our method.

- (a) The subject's behavior in circuit design is observed and the results are analyzed, using well-defined *attributes*.
- (b) The *domain knowledge "plus"* by means of testing in the broad sense of the word is investigated using latent trait models, and the attributes are related with the domain knowledge "plus".
- (c) The above two processes are repeated, with dynamic interactions between the two.

Here by an *attribute* we mean any behavior related with cognitive diagnosis, including an *episode*, a specific sequence of episodes, a behavior pattern representing a *buggy strategy*, etc. Basically, attributes are observable.

Two samples of subjects are needed in this approach. By Sample 1 we mean several hundred (or more) individuals representing the target population, for whom a specified cognitive diagnosis is considered. We need this sample for operationally defining the *competency space*, which represents domain knowledge "plus". The actual process will be initiated by administering the tests developed for this purpose. By Sample 2 we mean a smaller group of individuals sampled from the same population, who conduct actual, intensive tasks (e.g., circuit design) in the experimental situation.

Sample 2 can be a subgroup of Sample 1. If this is not the case, then these individuals in Sample 2 must take the computerized adaptive versions of the same tests so that their positions in the competency space can be estimated.

[II.2] Domain Knowledge "Plus"

Domain knowledge in circuit designing and electronic trouble-shooting includes understandings of *Boolean algebra*, *truth tables*, *Karnaugh maps* (K-maps), *logical gates*, their relationships with each other, etc., among others. They can be approached by means of tests. Also tests can deal with complete design tasks for *combinational circuits*, for example, especially if we use computerized tests. They are beyond the level of domain knowledge, and correspond to *plus*

in the subtitle.

[II.3] Preliminary Study: Two Quizzes

Two quizzes have been given to Vanderbilt sophomores of engineering, who were in one of Dr. Bharat Bhuva's courses. Quiz 1 has 4 questions and was given to 31 students, and Quiz 2 has 16 questions and was given to 24 students, a subset of the 31 who had taken Quiz 1.

In contents, Quiz 1 has two categories: (a) to generate a truth table for a specified gate (2 questions), and (b) to implement a specified 2-input gate using a specified type of gate only (2 questions); Quiz 2 has four categories to question equivalence of: (c) a K-map to a truth table (4 questions), (d) a K-map to a Boolean equation (4 questions), (e) a Boolean equation to a circuit (3 questions), and (f) two Boolean expressions (5 questions). The average proportions correct are: (a) 0.935, (b) 0.306, (c) 0.677, (d) 0.958, (e) 0.639 and (f) 0.833, respectively. With these small samples of 31 and 24 students, the resulting item score matrices look as if these tasks could be interpreted by one dimension, although some deviations are suggested in (e) and (f) of Quiz 2.

However, in operationally defining the competency space for the electronic trouble-shooting and circuit design, a larger dimensionality is expected.

[II.4] Different Types of Test Items

There are different types of test items, each of them has its own merits and demerits.

[II.4.1] PAPER-AND-PENCIL TESTS VS. COMPUTERIZED TESTS

Paper-and-pencil testing enables us to collect data within a limited amount of time for a large sample of examinees, whereas computerized testing requires a greater amount of time and also costs more. Computerized testing enables us to conduct more intensive research in a tractable environment, however, and to trace the examinee's behavior sequentially and in more detail, in comparison with paper-and-pencil testing.

It will be beneficial for our research, therefore, to use both and make the best use of their separate strengths. From practical aspects, a combination of 8:2, 7:3 or 6:4 of the paper-and-pencil and computerized test items may be desirable.

[II.4.2] RECOGNITION VS. CONSTRUCTION

A type of test item widely used for *recognition* is the multiple-choice test item, while a widely used type for *construction*, which can also be called *component design tasks* in the research of circuit design, is the open-ended test item.

In general, construction tasks may be more appropriate, but recognition tasks take less time for administration, and can be used effectively for certain questions, especially if we select

appropriate *distractors* in multiple-choice test items and also devise to discourage random or educated guessing. One such device may be to accomodate more than one correct answer without telling the examinees how many alternative answers are correct in a specific item, with such an instruction that: Find all expressions that are equivalent to <Given>.

Effective use of distractors in a multiple-choice test item may be exemplified by such a question that: <Given> $\overline{(x + y)}$ as a Boolean expression, a set of alternative answers in <Terminal> includes $\bar{x} + \bar{y}$, a common misconception, or *buggy DeMorgan*, in addition to the correct answer, $\bar{x}\bar{y}$. In this way, multiple-choice test items can be used effectively for detecting the adoption of *erroneous rules*. *Plausibility functions* of disctactors of multiple-choice test items (Samejima, 1984, 1994a) can be estimated using a nonparametric approach.

[II.4.3] PROBLEM COMPLEXITY

Various levels of problem complexity can be conceived of, including single-step problems as well as multi-step problems. Roughly speaking, paper-and-pencil testing can handle single-step problems and relatively simple multi-step problems without difficulty, while with computerized testing we can deal with more complex multi-step problems, tracing the examinee's cognitive processes and obtaining much more detailed information about his cognitive processes.

Take Question 14 of Quiz 2 as an example. In the quiz, the examinee is asked to decide if the Boolean equation

$$\bar{q}\bar{p} + q\bar{p} + p\bar{q} = \bar{p}\bar{q}$$

is true or false. To answer this question, perhaps the backward processes will be easier, which includes DeMorgan's law, tautology, distributive law, commutative law and absorption as shown below.

$$\begin{aligned}\bar{p}\bar{q} &= \bar{p} + \bar{q} && \text{DeMorgan's law} \\ &= \bar{p}(q + \bar{q}) + \bar{q}(p + \bar{p}) && \text{tautology} \\ &= \bar{p}q + \bar{p}\bar{q} + \bar{q}p + \bar{q}\bar{p} && \text{distributive law} \\ &= \bar{q}\bar{p} + \bar{q}\bar{p} + q\bar{p} + p\bar{q} && \text{commutative law} \\ &= \bar{q}\bar{p} + q\bar{p} + p\bar{q} && \text{absorption}\end{aligned}$$

This example belongs to the category of relatively simple multi-step problems. Note that not all steps are equally easy or difficult. In the above example, the use of *tautology* may be more difficult than the use of *commutative law*, for example.

[II.4.4] CONVENTIONAL TESTS VS. ADAPTIVE TESTS

Conventional tests are represented by paper-and-pencil tests and computerized tests, the latter of which have been rapidly put into practice in the past decade. Computerized adaptive tests have also been materialized as advanced technologies have made them more and more feasible. A strength of adaptive testing is that only a *tailored* subset of the total set of items, or *itempool*, is administered to an individual subject, and yet the loss in accuracy of estimation of his ability can be small.

In the present methodology, first, conventional tests will be administered to Sample 1 and these results will lead to the definition of the competency space, including the discovery of its dimensionality, and also to the *item calibration* of the itempool. After these have been accomplished, we will switch to computerized adaptive tests, using the results of the item calibration. Each individual of Sample 2 will take the adaptive tests, unless he has already taken the original conventional tests.

A strength of the computerized adaptive test also exists in item calibration. It has been shown that *on-line* item calibration can be conducted just as accurately as the conventional item calibration, in spite of the fact that the number of test items given to an individual subject is much less (e.g., 15 vs. 50) than that of the conventional test, or itempool, and also the number of examinees for the computerized adaptive test is much less (e.g., 1,500 vs. 3,000) than that for the conventional test (see Samejima, 1988, 1990).

[II.5] Computerized Tests with <Given>, <Hint> and <Terminal>

If we insert <Hint> between <Given> and <Terminal>, then <Hint> will control the difficulty of the question. Two types of sequential presentations of <Hint>, both starting with *no* <Hint> in case the examinee will solve the problem without depending upon any given hints, are conceivable. In the example given in [II.4.3], the left-hand-side of the first line will become <Given>, and the last line will be <Terminal>, (or vice versa), and each line from the top can be presented as a hint, thus constructing a *sequence* of hints.

Another method is to start with a more difficult hint, and, if the examinee fails in supplementing the remaining processes, a more obvious hint will be given, and so on. In the previous example, suppose that <Hint> and <Terminal> are reversed. Then the first step will be to add $\bar{q}\bar{p}$ at the beginning of the expression in <Given>. The first hint may be

$$a + b = a + b + b \quad .$$

If the examinee cannot use the hint for solving the problem, then delete it, and present a stronger hint such as

$$a + b = a + a + b \quad .$$

If this still does not work, then we may replace it by

$$a + b + c = a + a + b + c \quad ,$$

which is more suggestive than the previous two. Inclusion of items of this type is important, and the *graded response model* (Samejima, 1969, 1972, 1994c) is readily applicable. To a lesser extent, we can also incorporate items of this type in paper-and-pencil tests.

There can be other paths to the <Terminal> besides those that go through given <Hint>. Does each path test the same rules? Nonparametric approaches for estimating the operating characteristics (Samejima, 1981, 1988, 1990, 1994d) will eventually discover the answer to this question.

[II.6] Diffused Attributes and Concentrated Attributes

In the nonparametric approach for estimating the operating characteristics (Samejima, 1981, 1988, 1990, 1994d), several ways of approximating the conditional distribution of ability, given its maximum likelihood estimate, have been introduced, using the method of moment for fitting a least squared polynomial to the set of maximum likelihood estimates of ability (cf. Samejima & Livingston, 1979).

Suppose that the abscissa of each of the two figures, which are presented in Figure 2-1, is a dimension of the competency space, and the ordinate indicates the maximum likelihood estimate of the construct represented by the abscissa. In these figures, the estimated position of each of the eight subjects is shown by an arrow, and the corresponding approximated conditional density function of the construct is drawn.

Let us assume, for example, there are three disjoint attributes, and they distribute among the eight subjects in the way that the areas under the curves of the conditional density functions are shaded in the left-hand-side figure. Then the proportioned marginal density functions of the three attributes will be as shown at the bottom of the figure. If this is the case, we shall say that the disjoint attributes are *concentrated* with respect to the competency dimension. When a single attribute has a proportioned marginal density function which is similar to the three proportioned marginal density functions in the figure, we will also say that the attribute is concentrated. Such a result suggests that this competency dimension is closely related with the attribute in question.

If, in contrast, the proportioned marginal density function of an attribute shapes flatly over a wide range of the dimension, as is illustrated in the right-hand-side figure, we shall say that the attribute is *diffused* with respect to the competency dimension. A couple of possible interpretations of such a result may be:

1. the competency dimension has nothing to do with the attribute, and, if this happens to every dimension, then a larger dimensionality of the competency space will be

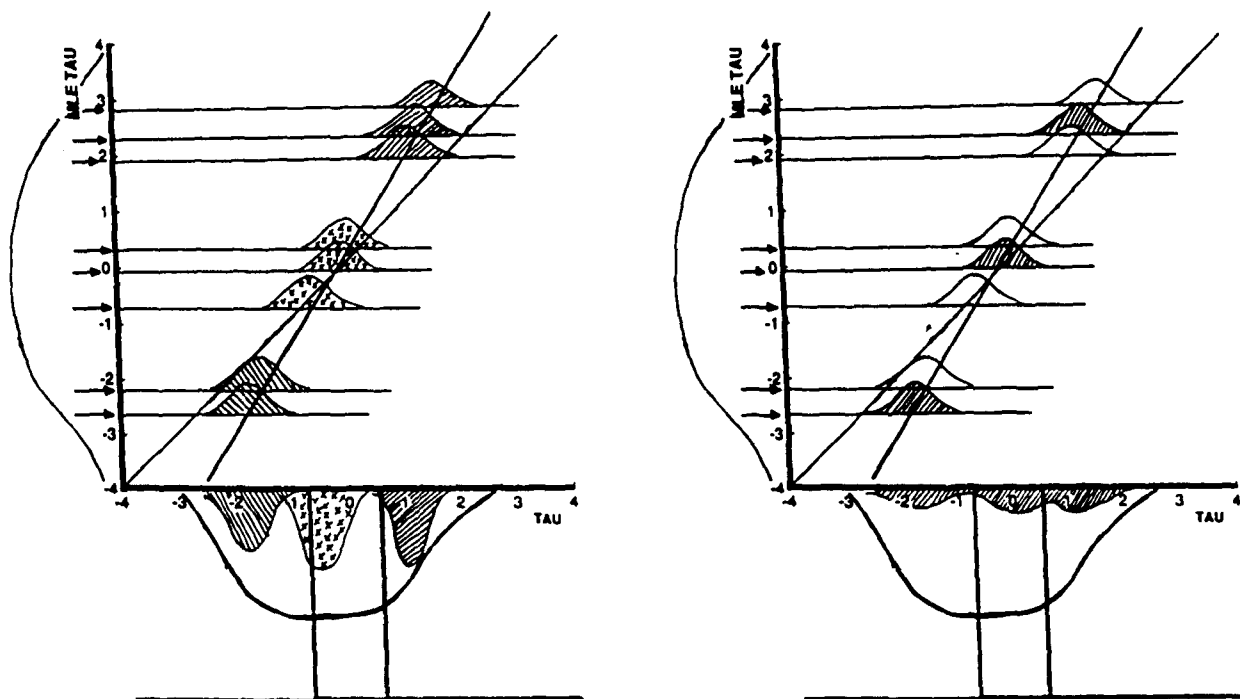


FIGURE 2-1

Schematic representation of a *concentrated* attribute (left) and a *diffused* attribute (right). The abscissa (τ) is a competency dimension, and the ordinate ($m\tau$) is the maximum likelihood estimate of the competency level of individuals.

needed, and

2. the attribute has multiple meanings.

Suppose that the attribute is a pattern of behavior. If it turns out to be concentrated, then we should investigate, by interviewing these subjects, etc., whether the attribute represents a specific family of strategies which is prone to be taken for subjects whose positions on this competency dimension are in the concentrated range. If this has been confirmed to be true, then in the future we should anticipate that a strategy in this family will be taken with a high probability by individuals in this range of the competency dimension, i.e., a finding of the research. If it turns out to be diffused with every dimension of our competency space, then we should investigate if the pattern of behavior has multiple meanings, that is, if it commonly belongs to separate families of strategies. If the answer is negative, then we must investigate the possibility of developing an additional set of test items to enhance the dimensionality of the competency space.

As we increase the size of Sample 2, we shall be able to estimate the operating characteristic of a specific attribute more and more accurately, using a nonparametric approach for estimating the operating characteristic of any discrete response, such as the Conditional P.D.F. Approach, which is described and discussed in Section 3.

[II.7] Alternative Method: Multi-Stage Latent Trait Approach

An alternative method for using Samples 1 and 2 is to combine them into one sample, which includes several hundred to one thousand individuals. In this method, first we need to develop software for, say, problem solving and/or designing tasks, after intensive pilot studies. With today's computer technologies and availabilities, it is possible to administer a set of, say, 30 items for several hundred individuals within a couple of months, if 10 to 15 carry-on microcomputers and the same number of testers are available. This sample size is comparable to typical sample sizes when paper-and-pencil tests are used in college environments.

Advancement of computer technologies has made it possible to use *figural responses* in computerized experiments, by using a mouse. This is especially beneficial to circuit design tasks on the gate level. An advantage of responses using a mouse also consists in the fact that the number of casual mistakes in responding will decrease, in comparison with responses using the key board.

This method includes technologies, which enable us to :

1. control an experimental situation by identical software accommodated in microcomputers,
2. have human subjects work on problem solving or designing tasks presented on their monitor screens,
3. have the microcomputers record their cognitive processes, and

4. have the computers analyze and evaluate the subjects' performances.

Thus methodologies originated in psychometrics could be adopted in cognitive psychology both in depth and perspective, which includes problem solving, trouble-shooting, etc.

If a sufficient research fund is available, this alternative method will be more fruitful. It is possible that the gate level trouble-shooting or circuit design tasks themselves can be incorporated into the software, making use of figural responses by a mouse. Thus dynamic interactions of microscopic and macroscopic approaches will be realized, enhancing the productivity of research.

Use of nonparametric approach (Samejima, 1981, 1988, 1990, 1994d) for *discovering* the meanings of patterns of behavior has an important role in the multi-stage latent trait approach. This is especially so in the presence of multi-strategies in problem solvings, multi-correct solutions, buggy strategies, and other factors that make understandings and evaluations of chunks of behavior difficult. Samejima (1984, 1994a) used a two-stage latent trait approach to discover the *plausibility function* of each distractor of each multiple-choice item of a vocabulary test. The multi-stage latent trait approach is similar in principle to this method.

In practice, in spite of complexities in understanding and evaluating chunks of behavior, it is likely that the performance of each individual can be dichotomously scored as *solution* and *nonsolution* with negligible ambiguity, as is the case with multiple-choice test items. Thus on the first stage of the multi-stage latent trait approach each problem or designing will be treated as a dichotomous item. Even if there are multi-strategies, if a single correct answer exists, the item should be scored either 0 or 1, ignoring different strategies. If multi-correct answers exist, in general, 1 should be given to all correct answers. This is a tentative treatment, and the separate operating characteristics for the separate multi-correct answers will be estimated later.

Factor analysis will be used to find out the dimensionality of the latent space. If more than one dimension are found, it will be wise to treat each dimension separately adopting unidimensional latent trait models as long as a *simple structure* exists, instead of turning to multidimensional latent trait models.

Some appropriate model for the dichotomous item, such as the normal ogive model or the logistic model, can be adopted for each latent dimension. Model validation for the adopted model will be made on the second stage.

A strength of the nonparametric approach developed by the principal investigator, which is introduced and discussed in Section 3, is that it can be used for relatively small sets of data with, say, several hundred to one thousand subjects. It is based on the *Old Test*, consisting of items whose characteristics are known. In the multi-stage latent trait approach the set of dichotomously scored items on the first stage can be used as the Old Test.

Based on this Old Test the operating characteristic of the *solution* will be estimated for model

validation. If the resulting curve of the estimated operating characteristic is close enough to the assumed parametric operating characteristic, then the adopted model is validated; if not, it is invalidated and the process must be repeated by adopting another dichotomous model.

After a model has been validated, using the nonparametric approach the operating characteristic of each chunk of behavior observed in each problem or designing can be estimated, and discoveries of their meanings will follow. For multi-correct solutions the operating characteristics can be estimated for separate correct solutions, and the results will clarify the order of the separate correct solutions.

The sets of resulting nonparametrically estimated operating characteristics may indicate which mathematical model is applicable. In multi-correct solution and/or multi-strategy cases a model developed for such cases (e.g., Samejima, 1983) will be necessary. Otherwise, they may direct us to the homogeneous case, and one of the models such as the normal ogive model and the logistic model may be appropriate; if they direct us to the heterogeneous case, then adoption of the acceleration model may be appropriate. In the latter case, a tentative parameterization of the nonparametrically estimated cumulative operating characteristics, using a very general semiparametric method (e.g., Ramsay & Wong, 1993) will be needed (see Samejima, 1994c).

[II.8] Grades of Attainment

It has been customary that diagnosis is made dichotomously, that is, individuals are categorized either in *mastery* or in *nonmastery* with respect to a given *attribute*, as exemplified by Tatsuoka's studies (Tatsuoka, 1985, 1990). Since each attribute involved in a task gets more and more complicated as mental processes get higher, however, *mastery* of an attribute requires a sequence of subprocesses. To give an example, consider the attribute, *fraction*, which is used by DiBello, Stout and Roussos (1993). They provide us with several items requiring this attribute, and one of them is item 3: "Solve: $7 - 2x = 9 + 3x$." It is noted that, in solving this problem, we need the understanding of the concept of fractions, but we do *not* need the *mastery* of the fraction, which includes addition, subtraction, multiplication and division of numerical and algebraic functions including fractions. If we *grade* the attainment for the fraction skills 0 (= no understandings), 1 (= understanding concept of fractions) and 2 (= mastery), instead of 0 (= non-mastery) and 1 (= mastery), then all we will need is grade 1 attainment in solving the above equation.

Thus in order to make an accurate cognitive diagnosis, introduction of the concept of grades of attainment (Samejima, 1994b) is advisable. This indicates that we need to turn to an appropriate graded response model, which is discussed in Section 4.

Grades of attainment are further discussed, together with comments on the DiBello-Stout diagnosis model (DiBello, Stout & Roussos, 1993) in Samejima, 1994b, and the reader is directed to this book chapter.

[II.9] Decomposition of the Competency Space

The competency space represents not only domain knowledge, but also *dynamics* of putting mastered chunks of knowledge in appropriate configurations was emphasized, which includes discovery of implications of what have been learned, integration and restructuring of the acquired domain knowledge, identification of necessary information in our long term memory, creation of new, innovative structures of the domain knowledge, etc. This is the reason why we say that the competency space represents domain knowledge *plus*.

Let Θ denote the total competency space, and be decomposed in such a way that

$$\Theta' = [\Theta'_a, \Theta'_b] ,$$

where Θ_a represents masteries of attributes and Θ_b consists of dimensions of dynamics which are beyond mastery of attributes (see Samejima, 1994b, 1994e). Thus diagnosis will be made in each of the two subspaces.

In cognitive diagnosis, this second subspace has rather been neglected. In some situations, however, diagnosis in Θ_b is more important, as exemplified by selection of Ph. D. candidates. There are graduate students who can do course work well, but are poor in designing dissertation research, for example. If diagnosis in Θ_b can be done well before decision of acceptance and rejection of applicants of a graduate program is made, this type of students will be screened and rejected.

For further details concerning this subject, the reader is directed to Samejima, 1994b.

References

- [1] DiBello, L. V., Stout, W. F. & Roussos, L. A. (1993). *Unified cognitive/psychometric diagnosis foundations and application*, Paper presented at the 1993 American Educational Research Association Meeting, April, Atlanta, GA.
- [2] Ramsay, J. O. & Wang, X. (1993). *Hybrid IRT models*. Paper presented at the meeting of the Psychometric Society, Berkeley, CA.
- [3] Samejima, F. (1969). Estimation of ability using a response pattern of graded scores, *Psychometrika Monograph, No. 17*.
- [4] Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph, No. 18*.
- [5] Samejima, F. (1981). *Final Report: Efficient methods of estimating the operating characteristics of item response categories and challenge to a new model for the multiple-choice item*. Knoxville, TN: Office of Naval Research Final Research Report of N00014-77-C-0360, Department of Psychology, University of Tennessee.
- [6] Samejima, F. (1983). *A latent trait model for differential strategies in cognitive processes*. Knoxville, TN: Office of Naval Research Report 83-1, N00014-81-C-0569, Department of Psychology, University of Tennessee.

- [7] Samejima, F. (1984). *Plausibility functions of Iowa Vocabulary test items estimated by the Simple Sum Procedure of the Conditional P.D.F. Approach*. Knoxville, TN: Office of Naval Research Report 84-1, N00014-81-C-0569, Department of Psychology, University of Tennessee.
- [8] Samejima, F. (1988). *Final Report: Advancement of latent trait theory*. Knoxville, TN: Office of Naval Research Final Research Report of N00014-81-C-0569, Department of Psychology, University of Tennessee.
- [9] Samejima, F. (1990). *Final Report: Validity study in multidimensional latent space and efficient computerized adaptive testing*. Knoxville, TN: Office of Naval Research Final Research Report of N00014-87-K-0320, Department of Psychology, University of Tennessee.
- [10] Samejima, F. (1991). *Usefulness of latent trait approach in cognitive diagnosis*. Paper presented at the ONR Conference on Cognitive Diagnosis. Pittsburg, PA; October, 1991.
- [11] Samejima, F. (1992). *A Design of Cognitive Diagnosis*. Paper presented at the ONR Conference on Cognitive Diagnosis/Model-Based Measurement. Champaign, IL; June, 1992.
- [12] Samejima, F. (1994a). Nonparametric estimation of the plausibility functions of the distractors of vocabulary test items. *Applied Psychological Measurement*, (in press).
- [13] Samejima, F. (1994b). A cognitive diagnosis method using latent trait models: competency space approach and its relationship with DiBello and Stout's unified cognitive/psychometric diagnosis model, in *Cognitive Diagnosis Assessment*, Hillsdale, NJ: Lawrence Erlbaum, (in press).
- [14] Samejima, F. (1994c). Acceleration model in the heterogeneous case of the general graded response model, *Psychometrika*, (in press).
- [15] Samejima, F. (1994d) Efficient nonparametric approaches for estimating the operating characteristics of discrete item responses. (accepted by *Psychometrika*).
- [16] Samejima, F. (1994e). *Cognitive diagnosis using latent trait models*. Paper presented in the symposium, *Approaches to Cognitive Modeling*, at the Annual Meeting of the National Council on Measurement in Education. New Orleans, LA; April, 1994.
- [17] Samejima, F. & Livingston, P. S. (1979). *Method of moments as the least squares solution for fitting a polynomial*. Knoxville, TN: Office of Naval Research Report 79-2, N00014-77-C-0360, Department of Psychology, University of Tennessee.
- [18] Tatsuoaka, K. K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics*, Vol. 12, 55-73.
- [19] Tatsuoaka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederickson, R. L. Glaser, A. M. Lesgold & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition*, pages 453-486. Hillsdale, NJ: Erlbaum.

III. Efficient Nonparametric Approaches for Estimating the Operating Characteristics of Discrete Item Responses

The principal investigator has been engaged in developing a family of approaches and methods for estimating the operating characteristic of a discrete item response, or the conditional probability, given latent trait, that the examinee's response be that specific response (Samejima, 1977b, 1981, 1988, 1990b). These methods are featured by the facts that:

1. estimation is made without assuming any mathematical forms, and
2. it is based upon a relatively small sample of several hundred to a few thousand examinees.

In this research period, rationale and the actual procedures of two nonparametric approaches, Bivariate P.D.F. Approach and Conditional P.D.F. Approach, the latter of which includes Simple Sum Procedure and Differential Weight Procedure, were integrated under the title, *Efficient nonparametric approaches for estimating the operating characteristics of discrete item responses*. In this paper, some examples of the results obtained by the Simple Sum Procedure and the Differential Weight Procedure of the Conditional P.D.F. Approach were given, using simulated data, and the usefulness of these nonparametric methods was also discussed. The paper was submitted to *Psychometrika*, and was accepted with minor modifications (see Section 1). Since modifications have not been made and it will take some time before it is published in *Psychometrika*, in this section, the outline of this paper is presented.

In estimating the operating characteristic of a discrete response, or the conditional probability, given ability, with which the discrete response occurs, there are two conceivable general approaches. One is the *parametric* approach, in which a specific mathematical model is assumed so that the estimation of the operating characteristic is reduced to the estimation of its item parameters. The other is the *nonparametric* approach, in which no mathematical model is involved, that is, estimation is made *without assuming any mathematical forms* for the operating characteristic. The usefulness of the nonparametric approach lies in the fact that they will allow researchers to *venture* in new areas by *discovering* the true shapes of the operating characteristics rather than a priori *molding* them into specific mathematical forms.

Lord has developed a nonparametric method to estimate the operating characteristic and applied it for SAT Verbal test items (Lord, 1970), and the results led him to conclude that Birnbaum's three-parameter logistic model (Birnbaum, 1968) fitted well to the nonparametrically estimated item characteristic curves of these items. Samejima proposed Normal Approximation Method (Samejima, 1977b), and then several other nonparametric methods (Samejima, 1981, 1988, 1990b). Levine developed a nonparametric method based upon the multilinear formula scoring theory (Levine, 1984). While Lord's method is focused upon a large set of data such as those available at the Educational Testing Service, for example, Samejima's and Levine's methods make use of a relatively small set of data collected for, say, several hundred to a few

thousand examinees. The latter methods can effectively be used for the *on-line item calibration* in computerized adaptive testing as well as for the item calibration in paper-and-pencil testing.

[III.1] Rationale

Let θ be *ability*, or latent trait, which takes on any real number. It is assumed that there is a set of test items measuring θ whose characteristics are known. This set of test items is called *Old Test*. Let $f(\theta)$ be the probability density function of θ , g denote a *target* test item for which the operating characteristics of the discrete responses are to be estimated, K_g be the discrete response to item g , and k_g denote a specific discrete response.

The joint density function, $\xi(k_g, \theta)$, of the discrete item response k_g to the target item g and ability θ is expressed as

$$\xi(k_g, \theta) = f(\theta) \text{prob.}[K_g = k_g | \theta] ,$$

which leads to

$$f(\theta) = \sum_{k_g} \xi(k_g, \theta) .$$

Thus the *operating characteristic*, $P_{k_g}(\theta)$, of the discrete item response k_g , or the conditional probability assigned to k_g , given θ , is provided by

$$P_{k_g}(\theta) \equiv \text{prob.}[K_g = k_g | \theta] = \frac{\xi(k_g, \theta)}{f(\theta)} = \frac{\xi(k_g, \theta)}{\sum_{i \in K_g} \xi(i, \theta)} . \quad (3.1)$$

Suppose that τ is a *one-to-one mapping* of θ which satisfies

$$\frac{d\theta}{d\tau} > 0 .$$

Then $P_{k_g}(\theta)$ can be written, analogously, as

$$P_{k_g}(\theta) = \text{prob.}[K_g = k_g | \tau] = \frac{\xi^*(k_g, \tau)}{f^*(\tau)} = \frac{\xi^*(k_g, \tau)}{\sum_{i \in K_g} \xi^*(i, \tau)} , \quad (3.2)$$

where $f^*(\tau)$ and $\xi^*(i, \tau)$ are the density function of the transformed ability τ and the joint density function of the discrete item response $i \in K_g$ and τ , respectively. Note the relationships

$$f^*(\tau) = f(\theta) \frac{d\theta}{d\tau}$$

and

$$\xi^*(k_g, \tau) = f^*(\tau) \text{prob.}[K_g = k_g | \tau] = f^*(\tau) P_{k_g}(\theta) = \xi(k_g, \theta) \frac{d\theta}{d\tau} ,$$

which are obtainable from the definition of τ , and (3.1) and (3.2). For simplicity, hereafter, k_g will be used for both a specific discrete item response to item g , and the event $K_g = k_g$. Similar usage of symbols will be made for certain other concepts and events.

Let $h (= 1, 2, \dots, n)$ denote an item of the Old Test, k_h be a discrete response to item h , and $P_{k_h}(\theta)$ denote the operating characteristic of k_h , or the conditional probability assigned to k_h , given θ . It is assumed that $P_{k_h}(\theta)$ is three-times differentiable with respect to θ . A response pattern based upon the Old Test, which is denoted by V , is given by

$$V = \{ K_h \}' ,$$

and its realization, v , can be written as

$$v = \{ k_h \}' .$$

Throughout the rest of this paper *local independence* (Lord & Novick, 1968) is assumed to hold, so that within any group of examinees all characterized by the same value of the latent variable θ , or its transformation τ , the distributions of the discrete item responses are all independent of each other. Let $\varphi^*(k_g, \tau, v)$ denote the tri-variate density function of k_g , τ and v . Thus (3.2) can be rewritten as

$$P_{k_g}(\theta) = \frac{\sum_v \varphi^*(k_g, \tau, v)}{\sum_{i \in K_g} \sum_v \varphi^*(i, \tau, v)} \quad (3.3)$$

There are many variations of the expression of the right-hand-side of (3.3), and one of them is

$$\begin{aligned} P_{k_g}(\theta) &= \frac{\sum_v \zeta^*(\tau | k_g, v) \text{prob.}[v \cap k_g]}{\sum_v \phi^*(\tau | v) \text{prob.}(v)} \\ &= \frac{\sum_v \zeta^*(\tau | k_g, v) \text{prob.}[v \cap k_g]}{\sum_{i \in K_g} \sum_v \text{prob.}[v \cap i]} , \end{aligned} \quad (3.4)$$

where $\phi^*(\tau | v)$ is the conditional density function of τ , given v , and $\zeta^*(\tau | k_g, v)$ is the conditional density function of τ , given k_g and v , which are provided by

$$\phi^*(\tau | v) = \frac{f^*(\tau) \text{prob.}[v | \tau]}{\text{prob.}(v)} \quad (3.5)$$

and

$$\zeta^*(\tau | k_g, v) = \frac{\xi^*(k_g, \tau) \text{prob.}[v | \tau]}{\text{prob.}[v \cap k_g]} \quad (3.6)$$

respectively.

Note that the joint density function, $\xi^*(k_g, \tau)$, of k_g and τ can be written, from (3.6), as

$$\xi^*(k_g, \tau) = \frac{\sum_v \zeta^*(\tau | k_g, v) \text{prob.}[v \cap k_g]}{\sum_v \text{prob.}[v | \tau]}$$

and then from this and (3.5) the density function of τ is given by

$$\begin{aligned} f^*(\tau) &= \frac{\sum_v \phi^*(\tau | v) \text{prob.}(v)}{\sum_v \text{prob.}[v | \tau]} \\ &= \frac{\sum_{i \in K_g} \sum_v \text{prob.}[v \cap i]}{\sum_v \text{prob.}[v | \tau]} . \end{aligned}$$

[III.2] Bivariate P.D.F. Approach

Direct approach to (3.3) or (3.4) is extremely difficult, for in so doing good estimates of $\varphi^*(i, \tau, v)$ or those of φ for all combinations of $i \in K_g$ and v , in addition to the set of $P_{k_h}(\theta)$'s for the n Old Test items. Some indirect approach is needed, therefore, to make use of (3.3) or (3.4).

The method called Bivariate P.D.F. Approach (Samejima, 1981) is an indirect approach based on (3.3), and in which p.d.f. stands for the probability density function of τ and its maximum likelihood estimate $\hat{\tau}$ obtained from the responses to the Old Test items. In this approach, the estimator of the operating characteristic is defined by

$$\hat{P}_{k_g}(\theta) \equiv \frac{\sum_{\hat{\tau}} \hat{\varphi}(k_g, \tau, \hat{\tau})}{\sum_{i \in K_g} \sum_{\hat{\tau}} \hat{\varphi}(i, \tau, \hat{\tau})} , \quad (3.7)$$

where $\varphi(k_g, \tau, \hat{\tau})$ is the tri-variate density function of k_g , τ and $\hat{\tau}$, $\hat{\varphi}(k_g, \tau, \hat{\tau})$ indicates the estimate of $\varphi(k_g, \tau, \hat{\tau})$ and $\sum_{\hat{\tau}}$ means the summation over all equally spaced values of $\hat{\tau}$ for which not all estimated bivariate densities, $\hat{\varphi}(k_g, \tau, \hat{\tau})$, are practically nil. It is noted that, in (3.7), $\hat{\tau}$ replaces the response pattern v in (3.3) and is treated as a continuous variable, and the ratio in the right-hand-side of (3.7) approximates the ratio of the integration of $\hat{\varphi}(k_g, \tau, \hat{\tau})$ with respect to $\hat{\tau}$ and the sum total of the integration of $\hat{\varphi}(i, \tau, \hat{\tau})$ over all $i \in K_g$. The question is how to estimate $\varphi(i, \tau, \hat{\tau})$ for all $i \in K_g$. To make it possible we need a specific transformation of θ to τ , which makes use of the test information function of the Old Test, and allows us to enjoy the benefit of mathematical simplicity.

The *item response information function* (Samejima, 1969, 1972) is defined by

$$I_{k_h}(\theta) \equiv -\frac{\partial^2}{\partial \theta^2} \log P_{k_h}(\theta) = \left\{ \frac{\frac{\partial}{\partial \theta} P_{k_h}(\theta)}{P_{k_h}(\theta)} \right\}^2 - \frac{\frac{\partial^2}{\partial \theta^2} P_{k_h}(\theta)}{P_{k_h}(\theta)} , \quad (3.8)$$

and the *item information function* is defined as the conditional expectation of $I_{k_h}(\theta)$ given θ , so that

$$I_h(\theta) \equiv E[I_{k_h}(\theta) | \theta] = \sum_{k_h} I_{k_h}(\theta) P_{k_h}(\theta) = \sum_{k_h} \frac{[\frac{\partial}{\partial \theta} P_{k_h}(\theta)]^2}{P_{k_h}(\theta)} . \quad (3.9)$$

Note that this item information function includes Birnbaum's item information function for the dichotomous test item (Birnbaum, 1968) as a special case. The operating characteristic, $P_v(\theta)$, of the response pattern v is defined as the conditional probability of v , given θ . Thus the operating characteristic of a given response pattern becomes the product of the operating characteristics of the item response categories contained in that response pattern, so that

$$P_v(\theta) = \prod_{k_h \in v} P_{k_h}(\theta) , \quad (3.10)$$

by virtue of local independence. The *response pattern information function*, $I_v(\theta)$, (Samejima, 1972) is given by

$$I_v(\theta) \equiv -\frac{\partial^2}{\partial \theta^2} \log P_v(\theta) = \sum_{k_h \in v} I_{k_h}(\theta) , \quad (3.11)$$

and the *test information function*, $I(\theta)$, is defined as the conditional expectation of $I_v(\theta)$, given θ , and from (3.8), (3.9), (3.10) and (3.11)

$$I(\theta) \equiv E[I_v(\theta) | \theta] = \sum_v I_v(\theta) P_v(\theta) = \sum_{h=1}^n I_h(\theta) .$$

The transformation of θ to τ is given by

$$\tau = \frac{1}{C_1} \int_{-\infty}^{\theta} [I(t)]^{1/2} dt + C_0 , \quad (3.12)$$

where C_0 is an arbitrary constant for adjusting the origin of τ , and C_1 is an arbitrary constant which equals the square root of the test information functions, $I^*(\tau)$, of τ was adopted. This transformation will be simplified and will become more manageable if a polynomial approximation to the square root of the test information function, $[I(\theta)]^{1/2}$, for the meaningful interval of θ is used, following the least squared errors principle. This can be accomplished by using the method of moments (see Samejima & Livingston, 1979). Thus (3.12) can be changed to the form

$$\begin{aligned} \tau &\doteq \frac{1}{C_1} \sum_{k=0}^m \frac{\alpha_k}{k+1} \theta^{k+1} + C_0 \\ &= \sum_{k=0}^{m+1} \alpha_k^* \theta^k , \end{aligned}$$

where α_k ($k = 0, 1, \dots, m$) is the k -th coefficient of the polynomial of degree m approximating

the square root of $I(\theta)$, and α_k^* is the k -th coefficient of the polynomial of degree $(m+1)$ transforming θ to τ , which is given by

$$\alpha_k^* \begin{cases} = C_0 & k = 0 \\ = \frac{\alpha_{k-1}}{C_1 k} & k = 1, 2, \dots, m+1 \end{cases} .$$

Adopting the maximum likelihood estimator, $\hat{\tau}_v$, as our estimator, its asymptotic normality with the two parameters, τ and $(1/C_1)$, is used as the approximation to the conditional distribution of $\hat{\tau}$, given τ (Samejima, 1977a). Note that the second parameter, which equals the standard deviation of the conditional distribution, is constant for all τ . By virtue of this constancy the first through fourth conditional moments of τ , given $\hat{\tau}$, can be obtained from the density function, $g(\hat{\tau})$, of $\hat{\tau}$ and the constant C_1 , by the following four formulas.

$$E(\tau | \hat{\tau}) = \hat{\tau} + \frac{1}{C_1^2} \frac{d}{d\hat{\tau}} \log g(\hat{\tau}) . \quad (3.13)$$

$$Var.(\tau | \hat{\tau}) = \frac{1}{C_1^2} \left\{ 1 + \frac{1}{C_1^2} \frac{d^2}{d\hat{\tau}^2} \log g(\hat{\tau}) \right\} . \quad (3.14)$$

$$E[\{\tau - E(\tau | \hat{\tau})\}^3 | \hat{\tau}] = \frac{1}{C_1^6} \left\{ \frac{d^3}{d\hat{\tau}^3} \log g(\hat{\tau}) \right\} . \quad (3.15)$$

$$\begin{aligned} E[\{\tau - E(\tau | \hat{\tau})\}^4 | \hat{\tau}] &= \frac{1}{C_1^4} \left\{ 3 + \frac{6}{C_1^2} \left[\frac{d^2}{d\hat{\tau}^2} \log g(\hat{\tau}) \right] + \frac{3}{C_1^4} \left[\frac{d^2}{d\hat{\tau}^2} \log g(\hat{\tau}) \right]^2 \right. \\ &\quad \left. + \frac{1}{C_1^4} \left[\frac{d^4}{d\hat{\tau}^4} \log g(\hat{\tau}) \right] \right\} . \end{aligned} \quad (3.16)$$

The first of these formulas is commonly seen in convolution transform (e.g., Hirschman & Widder), and used in statistical astronomy (Trumpler & Weaver, 1953), for example.

The marginal density function, $g(\hat{\tau})$, is not directly observable, but can be estimated from the set of the maximum likelihood estimates, $\hat{\tau}_s$'s, of the individual parameters τ_s 's, where $s (= 1, 2, \dots, N)$ denotes an individual examinee in our sample. This can be done by fitting a polynomial, following the least squared errors principle, using the method of moments (Samejima & Livingston, 1979).

The conditional density $\phi_{k_g}(\tau | \hat{\tau})$ can be estimated through these estimates of the conditional moments, by replacing $g(\hat{\tau})$ by $g_{k_g}(\hat{\tau})$, marginal density function for the subpopulation of examinees who share a specific response k_g to the target item g , in (3.13) through (3.16).

It will be appropriate to have the estimated conditional moments select a functional formula for $\hat{\phi}_{k_g}(\tau | \hat{\tau})$ for each equally spaced $\hat{\tau}$. One of the Pearson System distributions (e.g., Elderton and Johnson, 1969) will be selected for each of these conditional distributions, using the two coefficients, β_1 and β_2 , and Pearson's criterion κ , which can be written as

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} ,$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

and

$$\kappa = \frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)} .$$

In these three formulas, $Var.(\tau | \hat{\tau})$, $E[\{\tau - E(\tau | \hat{\tau})\}^3 | \hat{\tau}]$ and $E[\{\tau - E(\tau | \hat{\tau})\}^4 | \hat{\tau}]$ are substituted for μ_2 , μ_3 and μ_4 , respectively, which are obtained by formulas (3.14), (3.15) and (3.16), for the values of $\hat{\tau}$ which are appropriately selected with reasonably small, equal steps. If β_1 and β_2 turned out to be close to 0 and 3, respectively, then a normal density function may be used as the approximation to $\phi_{k_g}(\tau | \hat{\tau}_s)$. Otherwise, the criterion κ will lead to one of the Pearson System distributions, that is, if $\kappa < 0$, then Pearson's Type 1 distribution, which means the asymmetric β -distribution, will be selected, if $\kappa = 0$, $\beta_1 = 0$ and $\beta_2 < 3$, then Type 2 distribution, which is the symmetric β -distribution, will be assigned, if $0 < \kappa < 1$, then Type 4 distribution, if $\kappa > 1$, then Type 6 distribution, etc. Multiplying each Pearson System density function thus obtained by the estimated joint density function $\hat{g}_{k_g}(\hat{\tau}) (N_{k_g}/N)$, where N is the total number of examinees and N_{k_g} is the number of examinees who share the same response k_g to the target item g , $\hat{\varphi}(k_g, \tau, \hat{\tau})$ will be obtained.

It is noted that in estimating $\varphi(k_g, \tau, \hat{\tau})$ the Pearson System distributions are used, which are parametric. In this sense, the approaches that are introduced in the present paper are not strictly nonparametric, and the estimated conditional moments introduced earlier are used as the estimated parameters. In fact, if the interval $(-3.55, 3.55)$ is used with the step width 0.1 for $\hat{\tau}$ for a five-category response item, for example, then we are using as many as 1,420 estimated parameters for the single item.

Since this has to be done individually for each item g and for each and every discrete response k_g , and the process must be repeated as many times as the number of discrete item response categories for each and every item, it requires a substantial amount of CPU time. This is a drawback of this approach.

[III.3] Conditional P.D.F. Approach

In the Conditional P.D.F. Approach, several different procedures, which include Simple Sum Procedure, Weighted Sum Procedure, Proportioned Sum Procedure, Differential Weight Procedure, etc., have been considered (see Samejima, 1981). In this section, Simple Sum Approach and Differential Weight Approach will be introduced.

[III.3.1] SIMPLE SUM APPROACH

An estimator of $P_{k_g}(\theta)$ is defined as

$$\begin{aligned}\hat{P}_{k_g}(\theta) &\equiv \frac{\frac{1}{N} \sum_{s \in k_g} \phi^*(\tau | v_s)}{\frac{1}{N} \sum_{s=1}^N \phi^*(\tau | v_s)} \\ &= \frac{\sum_{s \in k_g} \phi^*(\tau | v_s)}{\sum_{s=1}^N \phi^*(\tau | v_s)}.\end{aligned}\quad (3.17)$$

This estimator does not require $f^*(\tau)$ nor $\text{prob.}[k_g | \tau]$, so it will be easy to use in practical situations. It is *not* a consistent estimator, however.

Proof: Consistency of the denominator:

Let N_v denote the number of examinees in our sample who share the same, specific response pattern v on the Old Test. Then

$$\begin{aligned}\frac{1}{N} \sum_{s=1}^N \phi^*(\tau | v_s) &= \frac{1}{N} \sum_v \sum_{s: v_s=v} \phi^*(\tau | v_s) = \sum_v \frac{N_v}{N} \phi^*(\tau | v) \\ &\rightarrow \sum_v \phi^*(\tau | v) \text{prob.}(v) .\end{aligned}$$

Thus it has been demonstrated that the denominator is consistent.

Proof: Inconsistency of the numerator:

Let $N_{v \cap k_g}$ be the number of examinees who share the same, specific response pattern v and the same, specific discrete response k_g to the target item g . From (3.5) and (3.6)

$$\varsigma^*(\tau | k_g, v) = \frac{\phi^*(\tau | v) \text{prob.}[k_g | \tau]}{\text{prob.}[k_g | v]} . \quad (3.18)$$

From this

$$\begin{aligned}\frac{1}{N} \sum_{s \in k_g} \phi^*(\tau | v_s) &= \frac{1}{N} \sum_v \sum_{s \in k_g: v_s=v} \phi^*(\tau | v_s) = \sum_v \frac{N_{v \cap k_g}}{N} \phi^*(\tau | v) \\ &\rightarrow \sum_v \phi^*(\tau | v) \text{prob.}(v \cap k_g) \\ &= \sum_v \varsigma^*(\tau | k_g, v) \text{prob.}[v \cap k_g] \frac{\text{prob.}[k_g | v]}{\text{prob.}[k_g | \tau]} .\end{aligned}\quad (3.19)$$

Thus the numerator is *not* consistent because of the *nuisance* factor shown as the ratio at the end of the last expression of (3.19).

Although this estimator is *inconsistent*, direct approach to (3.17) is possible, for it simply requires the set of $P_{k_h}(\theta)$'s for the n Old Test items. It has been named full information

simple sum formula, and tested, by Levine and Williams (Levine & Williams, 1991). The results show pretty good fits to the true operating characteristics, especially with large sample sizes.

In the Simple Sum Procedure of the Conditional P.D.F. Approach (Samejima, 1981, 1988, 1990b)

$$\hat{P}_{k_g}(\theta) \equiv \frac{\sum_{s \in k_g} \hat{\phi}(\tau | \hat{\tau}_s)}{\sum_{s=1}^N \hat{\phi}(\tau | \hat{\tau}_s)} \quad (3.20)$$

is adopted as our estimator of $P_{k_g}(\theta)$, which is almost identical with (3.17) except for the replacement of v_s by the maximum likelihood estimate $\hat{\tau}_s$ of the individual examinee s . Note that $\hat{\tau}_s$ is a function of v_s , but a one-to-one correspondence between $\hat{\tau}_s$ and v_s may not exist. If, for example, two Old Test items are *equivalent*, indicating that they share an identical set of operating characteristics, then two response patterns, in which the discrete responses to these two items are exchanged and the responses to all the other items are identical, will be distinct from each other, but will share the same $\hat{\tau}_s$. The lack of a one-to-one mapping between v_s and $\hat{\tau}_s$ will not affect the characteristics of the estimator (3.17) by this replacement, however. The estimated conditional density, $\hat{\phi}(\tau | \hat{\tau}_s)$, in (3.20) can be specified by using the estimated conditional moments of τ , given $\hat{\tau}_s$, which are given by (3.13) through (3.16), in a similar manner as in the Bivariate P.D.F. Approach, by substituting $\hat{\tau}_s$ for the equally spaced $\hat{\tau}$ in the Bivariate P.D.F. Approach. Note that this formula enables us to estimate all the operating characteristics of the discrete item responses for many different items almost simultaneously, which provides us with the benefit of economy in CPU time.

From both theory and practice, with many sets of data very high frequencies for the normal density function as approximations to $\phi(\tau | \hat{\tau}_s)$ are expected as the results of the above *branching*. When this is the case, Normal Approach Method (Samejima, 1981, 1988, 1990b), in which

$$\hat{\phi}(\tau | \hat{\tau}_s) = \frac{1}{[2\pi \text{Var.}(\tau | \hat{\tau}_s)]^{1/2}} \exp \left\{ \frac{(\tau - E(\tau | \hat{\tau}_s))^2}{2 \text{Var.}(\tau | \hat{\tau}_s)} \right\}$$

is adopted as the approximation to $\phi(\tau | \hat{\tau}_s)$, can be used. It can easily be seen that when $f^*(\tau)$ is normal this is approximately the case for all $\hat{\tau}_s$, and when it is uniform this is approximately the case for a wide range of $\hat{\tau}$.

This somewhat indirect simple sum approach includes several *smoothing* devices in the process by using polynomials obtained by the method of moments, etc. This makes the resulting estimated curves smoother than those obtained by the direct approach, a convenient feature when our sample size is relatively small. It has been shown that with simulated data this method provides us with the estimated operating characteristics which are very close to the *truth curves* (see Samejima, 1981, 1988).

The last factor in (3.19), the ratio of two conditional probabilities, enables us to make interesting observations. If our Old Test gives us a substantially large amount of information for the range of τ of interest, then the conditional distribution of τ , given v , becomes closer to a one point distribution at a specific value of τ , and the above ratio, or weighting factor, will become closer to unity at that value of τ . In such a case, consistency almost exists in the numerator of (3.17) and thus for the estimator itself. This means that the success of this method depends upon our choice of the Old Test, that is, whether the Old Test items satisfying the above condition is selected or not. When this is not the case, however, the estimated operating characteristics will have specific tendencies. If, for example, the *truth curve* is a steep, monotonically increasing one, then a substantially flatter estimated operating characteristic will be obtained, for the nuisance factor, that is, the ratio of $\text{prob.}[k_g | \hat{\tau}_s]$ to $\text{prob.}[k_g | \tau]$, will act as a smoothing factor.

It has been observed that Simple Sum Procedure combined with the Normal Approach Method works well especially in the *on-line item calibration of the adaptive testing* (see Samejima, 1981, 1988, 1990b). The reason is obvious from the above observation, since in the response pattern

$$v_s^* = \{v_s, k_g\}$$

where v_s is based upon a subtest of the itempool *tailored* for each individual examinee, and the conditional distribution of τ , given v_s , becomes closer to the one-point distribution at the true individual parameter, τ_s .

[III.3.2] DIFFERENTIAL WEIGHT APPROACH

When there already is a reasonably good estimate of $P_{k_g}(\theta)$ for each k_g , another approach, which in theory provides us with more accurate estimation, is possible. This approach is called Differential Weight Approach.

The *differential weight function*, $W_{k_g}(\tau; v)$, is defined by

$$W_{k_g}(\tau; v) \equiv \frac{\text{prob.}[k_g | \tau]}{\text{prob.}[k_g | v]} . \quad (3.21)$$

Then from this and (3.18)

$$\zeta^*(\tau | k_g, v) = \phi^*(\tau | v) W_{k_g}(\tau; v) . \quad (3.22)$$

Substituting (3.22) into (3.4), $P_{k_g}(\theta)$ can be written as

$$P_{k_g}(\theta) = \frac{\sum_v W_{k_g}(\tau; v) \phi^*(\tau | v) \text{prob.}[v \cap k_g]}{\sum_{k_g} \sum_v W_{k_g}(\tau; v) \phi^*(\tau | v) \text{prob.}[v \cap k_g]}$$

As before, let $s (= 1, 2, \dots, N)$ be a subject or an examinee in our sample. Define a *consistent* estimator of $P_{k_g}(\theta)$ by

$$\begin{aligned}\hat{P}_{k_g}(\theta) &\equiv \frac{\frac{1}{N} \sum_{s \in k_g} W_{k_g}(\tau; v_s) \phi^*(\tau | v_s)}{\frac{1}{N} \sum_{k_g} \sum_{s \in k_g} W_{k_g}(\tau; v_s) \phi^*(\tau | v_s)} \\ &= \frac{\sum_{s \in k_g} W_{k_g}(\tau; v_s) \phi^*(\tau | v_s)}{\sum_{k_g} \sum_{s \in k_g} W_{k_g}(\tau; v_s) \phi^*(\tau | v_s)} .\end{aligned}\quad (3.23)$$

Proof: Consistency of the numerator:

$$\begin{aligned}\frac{1}{N} \sum_{s \in k_g} W_{k_g}(\tau; v_s) \phi^*(\tau | v_s) &= \frac{1}{N} \sum_v \sum_{s \in k_g: v_s = v} W_{k_g}(\tau; v_s) \phi^*(\tau | v_s) \\ &= \sum_v \frac{N_{v \cap k_g}}{N} W_{k_g}(\tau; v) \phi^*(\tau | v) \\ &\rightarrow \sum_v W_{k_g}(\tau; v) \phi^*(\tau | v) \text{prob.}[v \cap k_g] .\end{aligned}\quad (3.24)$$

Thus it has been demonstrated that the numerator of (3.23) is consistent.

Proof: Consistency of the denominator:

From (3.24), straightforwardly

$$\frac{1}{N} \sum_{k_g} \sum_{s \in k_g} W_{k_g}(\tau; v_s) \phi^*(\tau | v_s) \rightarrow \sum_{k_g} \sum_v W_{k_g}(\tau; v) \phi^*(\tau | v) \text{prob.}[v \cap k_g] .$$

Therefore,

$$\hat{P}_{k_g}(\theta) \rightarrow P_{k_g}(\theta) .$$

Direct approach is possible if there already exists a reasonably good estimate of $\text{prob.}[k_g | \tau]$ or $P_{k_g}(\theta)$ itself, in addition to the set of $P_{k_h}(\theta)$'s for the n Old Test items. This can be accomplished by using the estimated $P_{k_g}(\theta)$ obtained by the full information simple sum formula.

Differential Weight Procedure of the Conditional P.D.F. Approach (Samejima, 1990a, 1990b) can be executed as a supplementary process of the Simple Sum Procedure combined with, say, the Normal Approach Method. For the estimator of $P_{k_g}(\theta)$,

$$\hat{P}_{k_g}(\theta) \equiv \frac{\sum_{s \in k_g} \hat{W}_{k_g}(\tau; \hat{\tau}_s) \hat{\phi}(\tau | \hat{\tau}_s)}{\sum_{s=1}^N \hat{W}_{k_g}(\tau; \hat{\tau}_s) \hat{\phi}(\tau | \hat{\tau}_s)} , \quad (3.25)$$

where $\hat{W}_{k_g}(\tau; \hat{\tau}_s)$ denotes the estimate of the *differential weight function* given by (3.21) by replacing v by $\hat{\tau}_s$. Note that this formula includes $\hat{\phi}(\tau | \hat{\tau}_s)$, but not $\hat{\zeta}(\tau | k_g, \hat{\tau}_s)$. Since $\hat{\phi}(\tau | \hat{\tau}_s)$ has already been obtained in the Simple Sum Procedure, all needed is to substitute the estimated operating characteristic of k_g obtained by the Simple Sum Procedure into the

specification of the estimate of the differential weight function $\hat{W}_{k_g}(\tau; \hat{\tau}_s)$. In so doing, it will be advisable to modify the estimated $P_{k_g}(\theta)$ obtained by the Simple Sum Procedure before using it in $\hat{W}_{k_g}(\tau; \hat{\tau}_s)$, if Old Test has a range of θ where the test information function $I(\theta)$ assumes low values. In many cases this happens on very high levels, or on very low levels, of θ , or both, relative to the ability distribution of our sample. In such a case, modifications can be made by extrapolating the portion of the estimated curve obtained in the interval of θ where the amount of test information provided by the Old Test is sufficiently large to the range of θ where this is not the case. Using the estimated $P_{k_g}(\theta)$ thus modified in $\hat{W}_{k_g}(\tau; \hat{\tau}_s)$, the reestimated operating characteristic will be obtained by (3.25).

To demonstrate how to use Simple Sum and Differential Weight Procedures of the Conditional P.D.F. Approach and to observe the results, part of a simulation study was introduced in the paper. The data were simulated data, provided by the Office of Naval Research as the initial itempool for the on-line item calibration research. There are one hundred hypothetical dichotomous test items in the itempool which were administered in the forms of conventional or non-adaptive tests. None of these one hundred items follow any specific mathematical models, and some of their item characteristic curves, or operating characteristics of the correct answer, are monotone increasing, but some others are not.

These one hundred items were divided into four subtests of twenty-five items each, which are called Subtests A, B, C and D. These subtests are combined into six pairs, that is, AB, AC, AD, BC, BD and CD of fifty items each. Six thousand hypothetical examinees were sampled from a population whose ability distribution is close to, but not quite equal to, $N(0, 1)$. One thousand hypothetical examinees were assigned to each of the six pairs of subtests. Thus each of the one hundred test items was administered to three thousand hypothetical examinees, and there were one thousand examinees who tried each pair of test items. The response pattern of each examinee was produced by the Monte Carlo method.

As for the details of the methods and the results of the comparisons, the reader is directed to Samejima, 1990a, or to the paper accepted by *Psychometrika* (see Section 1) when it is published.

References

- [1] Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. (Contributed chapters in Lord, F. M. & Novick, M. R., *Statistical theories of mental test scores, Chapters 17-20*, Reading, MA: Addison Wesley.)
- [2] Elderton, W. P. & Johnson, N. L. (1969). *Systems of frequency curves*. New York, N. Y.: Cambridge University Press.
- [3] Hirschman, I. I. & Widder, D. V. (1955). *The convolution transform*. Princeton, NJ: Princeton University Press.
- [4] Levine, M. V. (1984). *An introduction to multilinear formula score theory*. Champaign, IL: Office of Naval Research Report 84-4, N00014-83-K-0397, Model-Based Measurement Laboratory, Educational

Building, University of Illinois.

- [5] Levine, M. & Williams, B. (1991). *Nonparametric item response function estimation strategies*. Paper presented at the ONR Conference on Model-Based Measurement, Princeton, New Jersey.
- [6] Lord, F. M. (1970). Item characteristic curves estimated without knowledge of their mathematical form: A confrontation of Birnbaum's logistical model. *Psychometrika*, 35, 43-50.
- [7] Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison Wesley.
- [8] Samejima, F. (1969). Estimation of ability using a response pattern of graded scores, *Psychometrika Monograph*, No. 17.
- [9] Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph*, No. 18.
- [10] Samejima, F. (1977a). Effects of individual optimization in setting the boundaries of dichotomous items on the accuracy of estimation. *Applied Psychological Measurement*, 1, 77-94.
- [11] Samejima, F. (1977b). A method of estimating item characteristic functions using the maximum likelihood estimate of ability. *Psychometrika*, 42, 163-191.
- [12] Samejima, F. (1981). *Final Report: Efficient methods of estimating the operating characteristics of item response categories and challenge to a new model for the multiple-choice item*. Knoxville, TN: Office of Naval Research Final Research Report of N00014-77-C-0360, Department of Psychology, University of Tennessee.
- [13] Samejima, F. (1988). *Final Report: Advancement of latent trait theory*. Knoxville, TN: Office of Naval Research Final Research Report of N00014-81-C-0569, Department of Psychology, University of Tennessee.
- [14] Samejima, F. (1990a). *Differential Weight Procedure of the Conditional P.D.F. Approach for estimating the operating characteristics of discrete item responses*. Knoxville, TN: Office of Naval Research Report 90-4, N00014-87-K-0320, Department of Psychology, University of Tennessee.
- [15] Samejima, F. (1990b). *Final Report: Validity study in multidimensional latent space and efficient computerized adaptive testing*. Knoxville, TN: Office of Naval Research Final Research Report of N00014-87-K-0320, Department of Psychology, University of Tennessee.
- [16] Samejima, F. (1993). Human psychological behavior: approached by latent trait models. *Proceedings of the IEEE International Workshop on Neuro-Fuzzy Control, Muroran, Japan*, 26-33.
- [17] Samejima, F. & Livingston, P. S. (1979). *Method of moments as the least squares solution for fitting a polynomial*. Knoxville, TN: Office of Naval Research Report 79-2, N00014-77-C-0360, Department of Psychology, University of Tennessee.
- [18] Trumpler, R. J. & Weaver, H. F. (1953). *Statistical Astronomy*. Berkeley, CA: University of California Press.

IV. Acceleration Model

The competency space approach to cognitive assessment eventually requires a family of mathematical models which is appropriate for modeling cognitive processes, and is robust enough to continue to be useful as research goes further in depth and precision. To answer this necessity, a family of models, called *acceleration model*, has been proposed and discussed during this research period.

[IV.1] Processing Functions

Suppose that a cognitive process, like *problem solving*, contains a *finite* or *enumerable* number of steps. The graded item score x_g ($= 0, 1, \dots, m_g$) to (problem solving) item g is assigned to the individuals who have successfully completed up to the step x_g but *failed* to complete the step $(x_g + 1)$.

The *processing function*, $M_{x_g}(\theta)$, is defined as the *joint conditional probability* with which the individual of latent trait level θ completes the step x_g successfully, under the conditions that:

1. the individual's ability level is θ , and
2. the steps up to $(x_g - 1)$ have already been completed successfully.

It is assumed that $M_{x_g}(\theta)$ is *non-decreasing* in θ , and

$$M_{x_g}(\theta) \begin{cases} = 1 & \text{for } x_g = 0 & \text{(no step yet, or starting point)} \\ = 0 & \text{for } x_g = m_g + 1 & \text{(cannot be attained) ,} \end{cases}$$

for all θ , where $(m_g + 1)$ is the *hypothesized* graded item score adjacent to and above m_g .

The *fundamental* theoretical framework (Samejima, 1972) is given by

$$P_{x_g}(\theta) = \prod_{u \leq x_g} M_u(\theta) [1 - M_{(x_g+1)}(\theta)] , \quad (4.1)$$

where $P_{x_g}(\theta)$ is the operating characteristic of the item score x_g , that is,

$$P_{x_g}(\theta) \equiv \text{prob.}[X_g = x_g \mid \theta] .$$

The *cumulative operating characteristic*, $P_{x_g}^*(\theta)$, is the conditional probability with which the individual of latent trait θ completes the cognitive process successfully up to the step x_g , or further, so that it can also be expressed in terms of processing functions by

$$P_{x_g}^*(\theta) = \prod_{u \leq x_g} M_u(\theta) , \quad (4.2)$$

and from (4.1) and (4.2)

$$P_{x_g}(\theta) = P_{x_g}^*(\theta) - P_{(x_g+1)}^*(\theta) .$$

[IV.2] Criteria for Evaluating Graded Response or Partial Credit Models

Curve fitting and mathematical modeling are two different things. Even if a model fits data well, it cannot be an ultimate reason for accepting the model. Instead, the following five features have been considered as criteria for evaluating models.

1. The *principle* behind the model and the set of accompanied *assumptions* agree with the *psychological reality* in question. This is by far the most important criterion.
2. The model provides *additivity* in the operating characteristics of the item scores or degrees of attainment. Additivity holds if the operating characteristics belong to the *same* mathematical model under *finer recategorizations* and *combinings* of two or more categories together. This is the second most important criterion. Note that graded item scores or partial credits are more or less arbitrary, that is, it is a common practice to change the grades A, B, C, D, F to *Pass*, *Fail*, for example. Also, with the advancement of computer technologies, it is quite possible to obtain more abundant information from the individual's performance in *computerized* experiments as research is proceeded, and thus we need *finer recategorizations* of the whole cognitive process.
3. The model can be *naturally generalized to a continuous response model*. This criterion is a natural extension of *additivity*.
4. The model satisfies the *unique maximum condition* (Samejima, 1969, 1972). Satisfaction of this condition assures that the *likelihood function* of *any* response pattern consisting of such response categories has a *unique* local or terminal maximum.
5. The model provides the *ordered modal points* of the operating characteristics in accordance with the item scores.

Samejima (1972) distinguished the *homogeneous case* and the *heterogeneous case* of the graded response model. By the homogeneous case we mean a family of models in which the cumulative operating characteristics, $P_{x_g}^*(\theta)$'s, for $x_g = 1, 2, \dots, m_g$ are *identical* in shape, and these m_g functions are positioned *alongside* the abscissa in accordance with the item score x_g , whereas in the heterogeneous case not all $P_{x_g}^*(\theta)$'s are identical in shape (see Samejima, 1972).

It has been observed that models in the homogeneous case tend to satisfy the above criteria to a greater extent, whereas for those in the heterogeneous case fulfillment of these criteria is more difficult. For a model in the homogeneous case, if the principle behind the model and the set of accompanied assumptions are acceptable for the psychological reality in question, and if it satisfies the unique maximum condition, then it can be said to be an appropriate model for the following reasons.

1. *Additivity* of the operating characteristics *always holds*.

2. The model is naturally expanded to a *continuous* response model.
3. If the model satisfies the *unique maximum condition*, then:
 - (a) A strict orderliness among the modal points of $P_{x_g}(\theta)$'s holds.
 - (b) Satisfaction of the *unique maximum condition* (e.g., in the normal ogive and logistic models) also holds for *combined* categories and *more finely classified* categories.
 - (c) Satisfaction of the *unique maximum condition* also holds for the generalized *continuous* response model (Samejima, 1973).

For a model in the heterogeneous case, the same is not true, as is exemplified later. In spite of this handicap, models in the heterogeneous case tend to possess a greater variety in shapes of the operating characteristics $P_{x_g}(\theta)$'s . Thus search for a family of models in the heterogeneous case which satisfies the above five criteria just as well as those models in the homogeneous case is desirable.

[IV.3] General Acceleration Model and a Specific Model in Which the Logistic Function is Used

The *acceleration model* has been proposed as a model in the heterogeneous case developed with these considerations in mind. The processing function in the acceleration model is given by

$$M_{x_g}(\theta) = [\Psi_{x_g}(\theta)]^{\xi_{x_g}} .$$

where $\xi_{x_g} (> 0)$ is the step *acceleration* parameter, $\Psi_{x_g}(\theta)$ is a strictly increasing, five times differentiable function of θ with *zero* and *unity* as its two asymptotes, and provides the conditional ratio,

$$\frac{\Psi_{x_g}(\theta) \frac{\partial^2}{\partial \theta^2} \Psi_{x_g}(\theta)}{[\frac{\partial}{\partial \theta} \Psi_{x_g}(\theta)]^2} ,$$

given θ , which *decreases* with θ . In this model, the value of θ at which the *discrimination* power of $M_{x_g}(\theta)$ is maximal *increases* with ξ_{x_g} . It is assumed that the whole process leading to the solution of the problem consists of a *finite* number of *clusters*, each containing one or more *steps*, and *within* each cluster the parameters in $\Psi_{x_g}(\theta)$ common.

As a specific model,

$$\Psi_{x_g}(\theta) = \frac{1}{1 + \exp[-D \alpha_{x_g}(\theta - \beta_{x_g})]} ,$$

where $D = 1.7$, $\alpha_{x_g} (> 0)$ is the *discrimination* parameter, and β_{x_g} is the *location* parameters, has been used. It has been demonstrated that in this model:

1. *additivity* of the operating characteristics (criterion 2) *practically* holds;
2. a *continuous response model* can be obtained (criterion 3) as the limiting situation in which there are *infinitely many* subprocesses in each step;

3. the *unique maximum condition* (criterion 4) is satisfied;
4. *orderliness of the modal points* of the operating characteristics (criterion 5) *practically* holds, except for unusual cases where the unidimensionality of the steps should be questioned.

In contrast to the acceleration model, the partial credit model (Masters, 1982) and the generalized partial credit model (Muraki, 1992) do not have *additivity*, and thus are inappropriate as models for typical graded response situations. These models are versions of Bock's nominal response model (Bock, 1972), which is based on the individual *choice* behavior. Although they satisfy the unique maximum condition and the modal points of the operating characteristics are ordered in accordance with the graded item scores, or partial credits, lack of additivity, and of generalizability to continuous models, is detrimental as graded response models.

A strength of the acceleration model lies in the fact that, even if a researcher has started with an inappropriate model, it will be easy to switch to the acceleration model (Samejima, 1994). Figures 4-1 and 4-2 present the operating characteristics of 6 graded responses following the partial credit model and those following the acceleration model, respectively. It is obvious that these two sets of curves are practically indistinguishable. These two sets of curves also demonstrate that success in curve fitting is not sufficient in validating the model.

For further details of the acceleration model, the reader is directed to Samejima, 1994.

References

- [1] Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- [2] Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- [3] Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm, *Applied Psychological Measurement*, 16, 159-176.
- [4] Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores, *Psychometrika Monograph*, No. 17.
- [5] Samejima, F. (1972). A general model for free-response data, *Psychometrika Monograph*, No. 18.
- [6] Samejima, F. (1973). Homogeneous case of the continuous response model, *Psychometrika*, 38, 203-219.
- [7] Samejima, F. (1994). Acceleration model in the heterogeneous case of the general graded response model, *Psychometrika*, (in press).

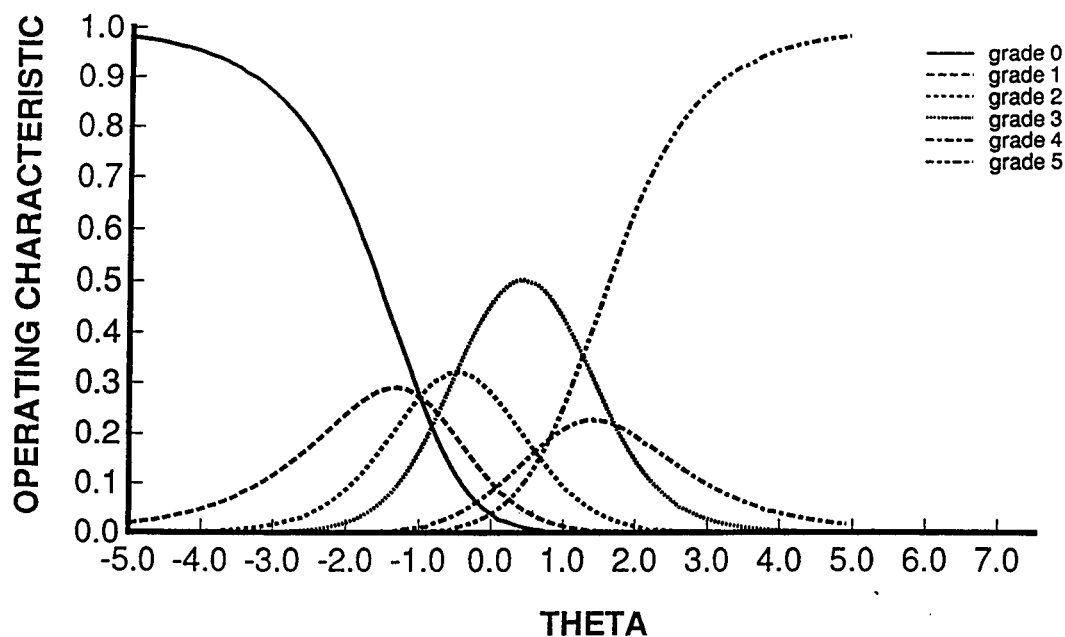


FIGURE 4-1

Six operating characteristics of graded item scores following the partial credit model.

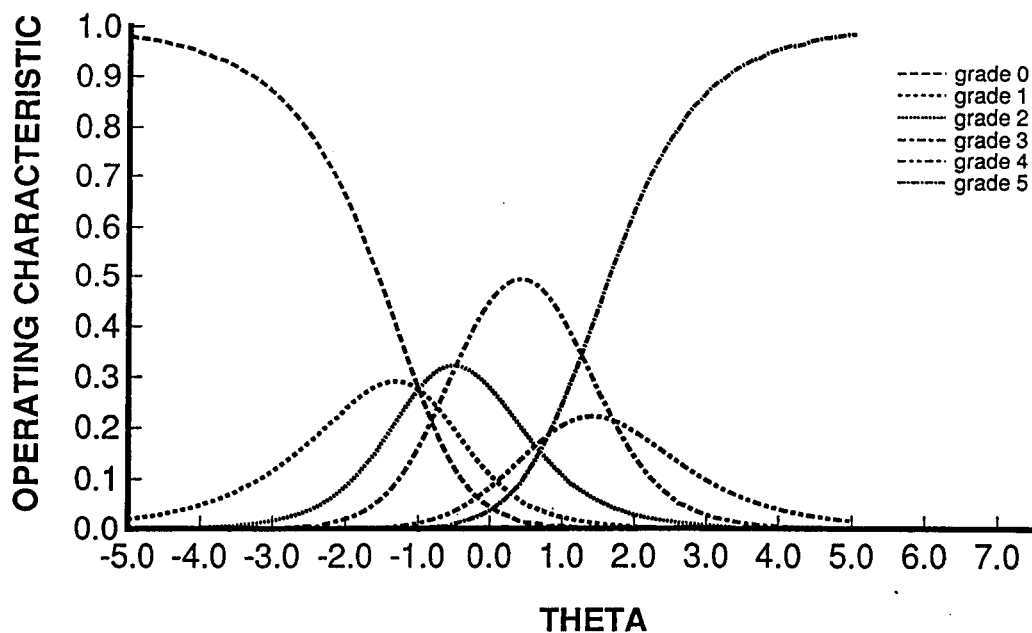


FIGURE 4-2

Six operating characteristics of graded item scores following the acceleration model. The parameters were adjusted so that the resulting operating characteristics be close to those in Figure 4-1.

V. Further Research and Integration of Research Findings

In this research period, some other topics that were worked on during the ONR funding years were further investigated and eventually published, or in press, in refereed journals, and also some of the research findings obtained during those years were integrated and published, or in press, as book chapters and a proceeding chapter. These topics are discussed below.

[V.1] Further Research

[V.1.1] MLE BIAS FUNCTION

Following the bias function of the maximum likelihood estimate in the three-parameter logistic model proposed by Lord (1983), the principal investigator expanded it for any discrete responses (Samejima, 1987), and called it the MLE bias function. The research was continued and eventually written in two articles shown as [4] and [5] of the refereed journal papers in Section 1 (pages 1 and 2), dividing the contents into the general case of discrete responses and a specific case of dichotomous responses.

[V.1.2] CRITICAL OBSERVATIONS OF THE TEST INFORMATION FUNCTION AS A MEASURE OF LOCAL ACCURACY

The principal investigator proposed the constant information model (Samejima, 1979a), and using this model observations were made with respect to the speed of convergence of the conditional distribution of the maximum likelihood estimate of ability, given its true value, as the number of items increases, to the asymptotic normality (Samejima, 1979b). Based on the results which indicated that there were substantial differences in the speed of convergence to the asymptotic normality depending on the fixed levels of ability, critical observations of the test information function as a measure of local accuracy in ability estimation were written in an article shown as [6] of the refereed journal papers in Section 1 (page 2).

[V.1.3] PLAUSIBILITY FUNCTIONS OF DISTRACTORS

Using the Simple Sum Procedure of the Conditional P.D.F. Approach (Samejima, 1981, 1988, 1990c), the operating characteristics of the distractors, called *plausibility functions*, of the multiple-choice test items of the Level 11 Vocabulary Subtest of the Iowa Tests of Basic Skill were estimated (Samejima, 1984). The results showed differential information from the separate distractors for most items, which can be used in ability estimation so that accuracies in estimation will be increased. These results were summarized and written in a paper under the title shown as [8] of the refereed journal papers in Section 1 (page 2).

[V.1.4] ESTIMATION OF RELIABILITY COEFFICIENTS USING THE TEST INFORMATION FUNCTION AND ITS MODIFICATIONS

While classical mental test theory is population-bound, latent trait models are population-free. The fact that the reliability coefficient of a test in classical mental test theory is a property of the population of individuals as well as of the test itself, it is still accepted as a magic number that solely belongs to the test. For different ability distributions, the reliability coefficients can be predicted (Samejima, 1990b), which clearly differ for different ability distributions. Predictions were made using the test information functions and also its two modifications (Samejima, 1990a), and the results were compared. These findings were written in a paper under the title shown as [9] of the refereed journal papers in Section 1 (page 2).

[V.2] Integration of Research Findings

[V.2.1] ROLES OF FISHER TYPE INFORMATION IN LATENT TRAIT MODELS

The roles of Fisher type information are important in latent trait models. They were integrated in a book chapter under the title shown as [1] of Section 1 (page 1). The topics include weakly parallel tests (Samejima, 1977), the test information function and its two modifications (Samejima, 1990a), predictions of the reliability coefficient and the standard error of estimation (Samejima, 1990b), equally discriminating ability scale (Samejima, 1981), nonparametric estimation of operating characteristics (Samejima, 1981, 1988, 1990c), constancy in the amount of information provided by a single dichotomous item (Samejima, 1979a), constant information model (Samejima, 1979a, 1979b), the MLE bias function (Samejima, 1987), among others.

[V.2.2] HUMAN PSYCHOLOGICAL BEHAVIOR

Human psychological behavior viewed from latent trait models was summarized and introduced to electronical engineering researchers working on neuro fuzzy control, and written as a proceedings chapter under the title shown as [11] of Section 1 (page 2). Among others, the paper includes the comprehensive methodologies for cognitive diagnosis using latent trait models, which were developed by the principal investigator.

[V.2.3] GRADED RESPONSE MODEL

The general theoretical framework of the graded response model and specific models such as the normal ogive and logistic model (Samejima, 1969, 1972), the partial credit model (Masters, 1982), the generalized partial credit model (Muraki, 1992), the acceleration model (Samejima, 1994), etc., were introduced and discussed in a book chapter under the title shown as [3] of Section 1 (page 1).

References

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

- [2] Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- [3] Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm, *Applied Psychological Measurement*, 16, 159-176.
- [4] Samejima, F. (1969). Estimation of ability using a response pattern of graded scores, *Psychometrika Monograph*, No. 17.
- [5] Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph*, No. 18.
- [6] Samejima, F. (1977). Weakly parallel tests in latent trait theory with some criticisms of classical test theory, *Psychometrika*, 42, 193-198.
- [7] Samejima, F. (1979a). *Constant information model: a new promising item characteristic function*. Knoxville, TN: Office of Naval Research Report, 79-1, N00014-77-C-0360, Department of Psychology, University of Tennessee.
- [8] Samejima, F. (1979b). *Convergence of the conditional distribution of the maximum likelihood estimate, given latent trait, to the asymptotic normality: observations made through the constant information model*. Knoxville, TN: Office of Naval Research Report, 79-3, N00014-77-C-0360, Department of Psychology, University of Tennessee.
- [9] Samejima, F. (1981). *Final Report: Efficient methods of estimating the operating characteristics of item response categories and challenge to a new model for the multiple-choice item*. Knoxville, TN: Office of Naval Research Final Research Report of N00014-77-C-0360, Department of Psychology, University of Tennessee.
- [10] Samejima, F. (1984). *Plausibility functions of Iowa Vocabulary test items estimated by the Simple Sum Procedure of the Conditional P.D.F. Approach*. Knoxville, TN: Office of Naval Research Report 84-1, N00014-81-C-0569, Department of Psychology, University of Tennessee.
- [11] Samejima, F. (1987). *Bias function of the maximum likelihood estimate of ability for discrete item responses*. Knoxville, TN: Office of Naval Research Report, 87-1, N00014-81-C-0569, Department of Psychology, University of Tennessee.
- [12] Samejima, F. (1988). *Final Report: Advancement of latent trait theory*. Knoxville, TN: Office of Naval Research Final Research Report of N00014-81-C-0569, Department of Psychology, University of Tennessee.
- [13] Samejima, F. (1990a.) *Modifications of the test information function*. Knoxville, TN: Office of Naval Research Report, 90-1, N00014-87-K-0320, Department of Psychology, University of Tennessee.
- [14] Samejima, F. (1990b). *Predictions of reliability coefficients and standard errors of measurement using the test information function and its modifications*. Knoxville, TN: Office of Naval Research Report, 90-2, N00014-87-K-0320, Department of Psychology, University of Tennessee.
- [15] Samejima, F. (1990c). *Differential Weight Procedure of the Conditional P.D.F. Approach for estimating the operating characteristics of discrete item responses*. Knoxville, TN: Office of Naval Research Report, 90-2, N00014-87-K-0320, Department of Psychology, University of Tennessee.
- [16] Samejima, F. (1994). Acceleration model in the heterogeneous case of the general graded response model, *Psychometrika*, (in press).

VI. Discussion

The author was too busy conducting research and writing research reports in her previous contract periods with the Office of Naval Research (N00014-77-C-0360, N00014-81-C-0569, N00014-87-K-0320) to write the research findings in the forms of refereed journal papers. It was her pleasure that in the second half of the present contract period she could write papers on her research outcomes obtained in the past years for refereed journals, book chapters, etc. Some of them were already published, and others are in press or accepted, as was described in Section 1.

Many other topics are still left unpublished in refereed journals, however, although they were printed in ONR research reports in the past years. They include the two topics mentioned in Section 1, that is, modified test information functions and the model for partly continuous and partly discrete responses. In addition to them, there are validity measures in latent trait models, a latent trait model for differential strategies, a family of models for multiple-choice test items, various outcomes from using the nonparametric approach to the estimation of the operating characteristic, computerized adaptive testing, and the practical usefulness of the method of moments for fitting polynomials collaborated with one of my former research assistants, Mr. Philip Livingston. Research will be supplemented on these topics, and they will eventually be published in refereed journals.

Many of these papers, published or unpublished, includes theories and methodologies, which will find their roles in cognitive diagnosis and assessment using controlled, computerized experiments with constructed responses, as was described in the previous sections of this final research report. Thus theory, methodologies and technologies necessary for cognitive assessment are ready for practical applications, provided that a sizable research fund is available to make the best use of advanced computer technologies.

It is the author's hope that the outcomes of the present research period, and of the previous ones started in 1977, will be used in the future, especially in cognitive diagnosis and assessment. The author believes that they will contribute to the advancement of psychology in depth as well as in perspectives.

Distribution List

Dr. Terry Ackerman 260 C Ed. Bldg. 1310 S. 6th St. University of Illinois Champaign, IL 61820-6990	Dr. Menucha Birenbaum School of Education Tel Aviv University Ramat Aviv 69978 ISRAEL	Dr. Raymond E. Christal UES LAMP Science Advisor AFHRL/MOEL Brooks AFB, TX 78235	Dr. Timothy Davey American College Testing Program P.O. Box 168 Iowa City, IA 52243
Dr. James Algina 1403 Norman Hall University of Florida Gainesville, FL 32605	Scientific Officer Code N712 Naval Training Systems Center Orlando, FL 32813-7100	Mr. Hua Hua Chung University of Illinois Department of Statistics 101 Illini Hall 725 South Wright St. Champaign, IL 61820	Dr. C. M. Dayton Department of Measurement Statistics & Evaluation College of Education University of Maryland College Park, MD 20742
Dr. Erling B. Andersen Department of Statistics Studiestraede 6 1455 Copenhagen DENMARK	Dr. Bruce Bloxom Defense Manpower Data Center 99 Pacific St. Suite 155A Monterey, CA 93943-3231	Dr. Norman Cliff Department of Psychology Univ. of So. California Los Angeles, CA 90089-1061	Dr. Ralph J. DeAyala EDMS-Benjamin Bldg., 1230 F University of Maryland College Park, MD 20742
Dr. Ronald Armstrong Rutgers University Graduate School of Management Newark, NJ 07102	Cdt. Arnold Bohrer Sectie Psychologisch Onderzoek Rekruterings-En Selectiecentrum Kwartier Koningen Astrid Bruijnstraat 1120 Brussels, BELGIUM	Director, Manpower Program Center for Naval Analyses 4401 Ford Avenue P.O. Box 16268 Alexandria, VA 22302-0268	Dr. Lou DiBello
Dr. Eva L. Baker 15366 Longbow Drive Sherman Oaks, CA 91403	Scientific Officer Code 281 Naval Training Systems Center Orlando, FL 32826-3224	Director, Manpower Support and Readiness Program Center for Naval Analysis 2000 North Beauregard Street Alexandria, VA 22311	Dr. Dattprasad Divgi Center for Naval Analysis 4401 Ford Avenue P.O. Box 16268 Alexandria, VA 22302-0268
Dr. Laura L. Barnes College of Education University of Toledo 2801 W. Bancroft Street Toledo, OH 43606	Dr. Robert Brennan American College Testing Programs P. O. Box 168 Iowa City, IA 52243	Dr. Stanley Collyer Office of Naval Technology Code 222 800 N. Quincy Street Arlington, VA 22217-5000	Dr. Hei-Ki Dong Bell Communications Research 6 Corporate Place PYA-1K226 Piscataway, NJ 08854
Dr. William M. Bart University of Minnesota Dept. of Educ. Psychology 330 Burton Hall 178 Pillsbury Dr., S.E. Minneapolis, MN 55455	Dr. John B. Carroll 409 Elliott Rd., North Chapel Hill, NC 27514	Dr. Hans F. Crombag Faculty of Law University of Limburg P.O. Box 616 Maastricht The NETHERLANDS 6200 MD	Dr. Fritz Drasgow University of Illinois Department of Psychology 603 E. Daniel St. Champaign, IL 61820
Dr. Isaac Bejar Mail Stop: 10-R Educational Testing Service Rosedale Road Princeton, NJ 08541	Dr. John M. Carroll IBM Watson Research Center User Interface Institute P.O. Box 704 Yorktown Heights, NY 10598	Ms. Carolyn R. Crone Johns Hopkins University Department of Psychology Charles & 34th Street Baltimore, MD 21218	Defense Technical Information Center Cameron Station, Bldg 5 Alexandria, VA 22314 (12 Copies)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Dr. Drew Gitomer
Educational Testing Service
Princeton, NJ 08541

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

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

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

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

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

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

Dr. Delwyn Hamisch
University of Illinois
51 Gerty Drive-#123
Champaign, IL 61820

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

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

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

Dr. Paul W. Holland

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

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

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

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

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

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

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

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

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

11/25/94

Dr. Brian Junker
232 Baker Hall
Carnegie Mellon University
Pittsburg, PA 15213

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

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

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

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

Mr. Soon-Hoon Kim
Computer-based Education
Research Laboratory
University of Illinois
Urbana, IL 61801

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

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

Dr. Richard J. Koubek

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

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

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

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

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

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

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

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

Library, NAWC-TSD
12350 Research Parkway
Orlando, FL 32826-3224

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

Dr. Richard Luecht

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

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

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

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

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

Mr. Christopher McCusker
University of Illinois
Department of Psychology
603 E. Daniel St.

Dr. Robert McKinley
Law Services
Box 40
Newtown, PA 18940-0040

Dr. Wim J. van der Linden
Universiteit Twente
P. O. Box 217
7500 AE Enschede
The Netherlands

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

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

Dr. Robert Mislav
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 MP1-20
Washington, DC 20380

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

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

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

Dr. Merry Bullock
Program Director
Human Cognition & Perception
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230

Dr. Harold F. O'Neill, Jr. School of Education - WPH 801 Department of Educational Psychology & Technology University of Southern California Los Angeles, CA 90089-0031	Dr. Mark D. Reckase ACT P. O. Box 168 Iowa City, IA 52243	Dr. Robin Shealy 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
Dr. James B. Olsen WICAT Systems 1875 South State Street Orem, UT 84058	Dr. Malcolme Ree AFHRL/MOA Brooks AFB, TX 78235	Dr. Kazuo Shigemasa 7-9-24 Kugenuma-Kaigan Fujisawa 251 JAPAN	Dr. Tomoichi Ishisuka The National Center for University Entrance Examination 2-19-23 Komaba, Meguroku Tokyo 153, Japan
Office of Naval Research, Code 1142CS 800 N. Quincy Street Arlington, VA 22217-5000 (6 Copies)	Mr. Steve Reiss N660 Elliott Hall University of Minnesota 75 E. River Road Minneapolis, MN 55455-0344	Dr. Randall Shumaker Naval Research Laboratory Code 5510 4555 Overlook Avenue, S.W. Washington, DC 20375-5000	Dr. John Tangney AFOSR/NL, Bldg. 410 Bolling AFB, DC 20332-6448
Dr. Judith Orasanu Basic Research Office Army Research Institute 5001 Eisenhower Avenue Alexandria, VA 22333	Dr. Carl Ross CNET-PDCD Building 90 Great Lakes NTC, IL 60088	Dr. Richard E. Snow School of Education Stanford University Stanford, CA 94305	Dr. Kikumi Tatsuoka Educational Testing Service Mail Stop 03-T Princeton, NJ 08541
Dr. Jesse Orlansky Institute for Defense Analyses 1801 N. Beauregard St. Alexandria, VA 22311	Dr. Fumiko Samejima Department of Psychology University of Tennessee 310B Austin Peay Bldg. Knoxville, TN 37916-0900	Dr. Richard C. Sorensen Navy Personnel R&D Center San Diego, CA 92152-6800	Dr. Maurice Tatsuoka 220 Education Bldg 1310 S. Sixth St. Champaign, IL 61820
Dr. Peter J. Pashley Educational Testing Service Rosedale Road Princeton, NJ 08541	Mr. Drew Sands NPRDC Code 62 San Diego, CA 92152-6800	Dr. Judy Spray ACT P.O. Box 168 Iowa City, IA 52243	Dr. David Thissen L. L. Thurstone Psychometric Laboratory Davie Hall University of North Carolina Chapel Hill, NC 27599
Wayne M. Patience American Council on Education GED Testing Service, Suite 20 One Dupont Circle, NW Washington, DC 20036	Lowell Schoer Psychological & Quantitative Foundations College of Education University of Iowa Iowa City, IA 52242	Dr. Peter Stolfoff Center for Naval Analysis 4401 Ford Avenue P.O. Box 16268 Alexandria, VA 22302-0268	Mr. Thomas J. Thomas Federal Express Corp., HRD 3035 Directors Row, Ste. 501 Memphis, TN 38131
Dr. James Paulson Department of Psychology Portland State University P.O. Box 751 Portland, OR 97207	Dr. Mary Schratz 905 Orchid Way Carlsbad, CA 92009	Dr. William Stout University of Illinois Department of Statistics 101 Illini Hall 725 South Wright St. Champaign, IL 61820	Dr. Gary Thomasson Defense Manpower Data Center 99 Pacific Street, Suite 155A Monterey, CA 93943-3231
Dept. of Administrative Sciences Code 54 Naval Postgraduate School Monterey, CA 93943-5026	Dr. Dan Segall Navy Personnel R&D Center San Diego, CA 92152		Dr. Robert Tsutakawa University of Missouri Department of Statistics 222 Math. Sciences Bldg. Columbia, MO 65211

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

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

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

Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

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

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

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

Dr. Brian Waters
HumRRO
1100 S. Washington
Alexandria, VA 22314

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

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

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

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

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

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

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

Dr. Hilda Wing
Federal Aviation Administration
800 Independence Ave, SW
Washington, DC 20591

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

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

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

Dr. Eiji Muraki
Educational Testing Service
Princeton, NJ 08541

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

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

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

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

Dr. Henry Braun
Vice President for Research
Educational Testing Service
Princeton, NJ 08541

Dr. Robert Brandewie
99 Pacific Street, Suite 155A
Defense Manpower Data Center
Personnel Testing Division
Monterey, CA 93940-2453

Dr. Merry Bullock
Program Director
Human Cognition & Perception
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230

Dr. Susan E. Chipman
Office of Naval Research
Code 342CS
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Cheryl E. Eavey
Program Director
Methodology, Measurement
& Statistics
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230

Ms. C. C. Everley
Department of Navy
Office of Naval Research
Resident Representative
101 Marietta Tower, Suite 2805
Atlanta, GA 30303

Dr. Irvin Katz
Educational Testing Service
Princeton, NJ 08541

Dr. Paul Nichols
765 Enderis Hall
P. O. Box 413
Department of Educational Psychology
University of Wisconsin-Milwaukee
Milwaukee, WI 53201

Dr. Shizuhiko Nishisato
The Ontario Institute for
Studies in Education
252 Bloor Street West
Toronto, Canada M5S 1V6

Dr. James O. Ramsay
Department of Psychology
Stewart Biological Sciences Bldg.
1205 Dr. Penfield Avenue
Montreal, QC, CA

Dr. Takahiro Sato
NEC C&C System Research Laboratories
4-1-1 Miyazaki, Miyamae-ku
Kawasaki, Kanagawaken 213
Japan

Dr. Sukeyori Shiba
1-8-3 Minami-Ogino
Suginamiku, Tokyo 167
Japan

Dr. Norio Suzuki
The National Center for
University Entrance Examination
2-19-23 Komaba, Meguroku
Tokyo 153, Japan

Dr. Bradford Symptom
Analytic Systems
P.O. Box 872
El Toro, CA 92630-0872

Dr. Haruo Yanai
The National Center for
University Entrance Examination
2-19-23 Komaba, Meguroku
Tokyo 153, Japan