This report explores an approach to item development and psychometric modeling which explicitly incorporates knowledge about the mental models used by examinees in the solution of items into a psychometric model that characterize performances on a test, as well as incorporating that knowledge into the item development process. The paper focuses on the hidden figure item type. Although there is an extensive literature on the correlates of performance for this type of item little is known about the mental models that may explain performance on the item. The approach taken in this paper is to search for a complexity dimension that accounts for the difficulty of hidden figures. Although several complexity dimensions can be postulated the one chosen was inspired by artificial intelligence research on vision. A computer-based system was developed to analyze as well as generate items based on this framework. To empirically determine the validity of the chosen framework two experiments were conducted. The results suggest that this approach to psychometric modeling is viable. The practical and theoretical implications of the research are discussed. (Author/LMO)
A Generative Approach to the Development of Hidden-Figure Items

Isaac I. Bejar
Educational Testing Service
Princeton, New Jersey

Peter Yocom
Rensselaer Polytechnic Institute
Troy, New York

June 1986

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

This research was sponsored by the Personnel and Training Research Program, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-83-C-0761, Contract Authority Identification No. NR 150-531.
A Generative Approach to the Development of Hidden-Figure Items

Isaac I. Bejer and Peter Yocam

This report explores an approach to item development and psychometric modeling which explicitly incorporates knowledge about the mental models used by examinees in the solution of items into a psychometric model that characterize performances on a test, as well as incorporating that knowledge into the item development process. The paper focuses on the hidden figure item type. Although there is an extensive literature on the correlates of performance for this type of item little is known about the mental models that may explain performance on the item. The approach taken in this paper is to search for a complexity dimension that accounts for the difficulty of hidden figures. Although several complexity dimensions can be postulated we chose one inspired by artificial intelligence research on vision. A computer-based system was developed to analyze as well as generate items based on this framework. To empirically determine the validity of the chosen framework two experiment were conducted. The results suggest that this approach to psychometric modeling is viable. The practical and theoretical implications of the research are discussed.
Abstract

This report explores an approach to item development and psychometric modeling which explicitly incorporates knowledge about the mental models used by examinees in the solution of items into a psychometric model that characterize performances on a test, as well as incorporating that knowledge into the item development process. The paper focuses on the hidden figure item type. Although there is an extensive literature on the correlates of performance for this type of item little is known about the mental models that may explain performance on the item. The approach taken in this paper is to search for a complexity dimension that accounts for the difficulty of hidden figures. Although several complexity dimensions can be postulated we chose one inspired by artificial intelligence research on vision. A computer-based system was developed to analyze as well as generate items based on this framework. To empirically determine the validity of the chosen framework two experiments were conducted. The results suggest that this approach to psychometric modeling is viable. The practical and theoretical implications of the research are discussed.
Acknowledgements

We would like to express our appreciation to Lisa Jansen for assisting with data collection and database maintenance, to Ka-Ling Chan for the statistical analysis and to Jessie Cryer for typing the manuscript as well as keeping track of everything. I’m also grateful to Susan Embretson and Phil Oltman for valuable comments. I’m especially grateful to Miguel A. García-Pérez for a number of insightful comments.
Test validation has traditionally focused on an accounting of response consistency. Indeed, the most comprehensive form of test validation, construct validity, has been described as implying "a joint convergent and discriminant strategy entailing both substantive coverage and response consistency in concert." (Messick, 1981, p. 575). There has been far less emphasis on an accounting of response effort (but see Campbell, 1961; Carroll, 1980; Davies & Davies, 1965; Egan, 1979; Elithorn, Jones, Kerr, & Lee, 1964; Tate, 1948; Zimmerman, 1954). These two focuses, response consistency and response effort, are not antithetical, by any means, see e.g., Embretson’s (1983) discussion of construct representation versus nonmothetic span. In fact an argument could be made, although it will not be elaborated here, that construct validity, in addition to requiring an accounting of substantive coverage and response consistency, also requires an accounting of response effort. That is, knowing the latent structure of a test—for example, its factorial structure or its fit to a particular item response model—is clearly essential to an interpretation of test scores but is not the entire story. An accounting of response effort would clearly enhance the validational status of a test because to obtain that accounting it is likely that a model incorporating the mental structures and processes needed to solve the item would be required. If this model has been previously and independently validated then, clearly, the validational status of the test will be enhanced.
Not only are accountings of response effort and consistency not antithetical, they entail almost parallel considerations. For example, within the response-consistency tradition, the extent to which covariation is accounted for by relevant and irrelevant (e.g., method) variables is often the basic data from which validity is assessed (e.g., Campbell & Fiske, 1959). Within the response-effort framework the contributions of relevant and irrelevant processes to difficulty could be similarly viewed. For example, patterns may have been inadvertently included in a test could affect the difficulty of items by cuing specifically coached test takers to the correct alternative. Within the response-consistency framework, when that occurs we say that examinees are not responding in accordance with their ability. This response behavior is in turn reflected in lack of fit of the item-response model. Within the response-effort framework we would say that examinees are not responding in accordance with the mental model postulated for a specific item and this response behavior would be manifested as a discrepancy between the estimated difficulty of the item, based on some item-response model, and the expected difficulty given the mental model for that item. Discrepancy between difficulty estimates are well entrenched in psychometrics. What may be new here is that one of the estimates is based on a substantive model of the effort required by an item. By contrast, in typical applications, for example, differential item performance, discrepancies in the difficulty estimates from different groups constitute the data.

An emphasis in accounting for response consistency is compatible with the latent trait approach to individual differences. This approach includes both factor analysis and item response theory (Lord, 1980). An accounting of response effort also fits well within item response theory but in
addition requires inspiration from cognitive science to formulate mental models of the item solution process. To see these two sets of considerations in action, consider a test for which we have established that some item response model fits perfectly. Moreover, through correlational analysis we have established that it is a "verbal" test. It is tempting to stop there and argue that the test has been validated. Indeed, many validation efforts stop at this point. There is, however, quite a bit more to explain. The items in the test differ in difficulty; some are very easy, others are very hard. This variation presents no major problem since every item response model includes a difficulty parameter. However, estimating the difficulties is not the same thing as explaining them. As a result we do not have a method, when it is time to create a new form of the test, to predict the psychometric characteristics of an item. The standard procedure followed by major testing organizations is to write many items and pretest them with the hope that enough of the items will survive the process and a new form can be constructed that resembles the previous one. This procedure is very effective, but it also underscores the fact that our understanding of the test is far from complete, for if it were, we should be able, for example, to construct forms that are parallel both substantially and psychometrically on an a priori basis.

The objective of this paper is to illustrate an approach to test modeling that encompasses both response consistency and response effort. We call this approach generative for two reasons. The approach is generative in the usual dictionary sense of the word—i.e., of "having the power of generating, originating, producing or reproducing"—in this case items with known psychometric characteristics. But the approach may be interpreted more broadly, as in the sense of Chomskyan linguistics in which a generative
grammar is defined as being capable of assigning a description to every sentence in the language and also capable of generating all the sentences in the language. The search for this type of grammar is a major preoccupation of some linguists.

A generative psychometrics, then, involves a "grammar" capable of assigning a psychometric description to every item in the universe of items and is also capable of generating all the items in the universe of items. Some of these ideas are implicit in certain item-generation schemes (e.g., Bormuth, 1970). However, the emphasis of these schemes was almost totally on generating rather than on assigning a description with psychometric utility to the generated items. In that sense, therefore, those approaches were incomplete.

Nothing in the definition proposed above dictates what sort of "description" should be attached to an item other than its psychometric utility. In the context of ability testing, it would be natural to assign a description with reference to an item-response model, or with reference to the response-time distribution. In a context of diagnostic testing the description might be with respect to a set of misconceptions, as in Brown & Burton (1978), Burton (1982); and see also Bejar (1984).

**Overview**

In this paper we are concerned with spatial ability, and therefore we will be concerned with a description of the item that has reference to both its difficulty and its response-time distribution. More concretely, this paper focuses on the hidden figure item type. Figure 1 shows two sample items. The role of the examinee is to determine whether the smaller figure is embedded in the larger figure.
This item type has been used extensively in field dependence-independence work; as a result there is an ample literature on correlating performance on hidden-figure tests with personality variables (e.g., Witkin, Goodenough, & Oltman, 1979). Unfortunately, nothing in that literature could be used as a means of constructing the grammar through which items could be generated and a description assigned. The grammar ultimately chosen for this item was inspired by artificial intelligence research in vision (Mayhew & Frisby, 1984) and is based on a pattern-recognition algorithm called the Hough transform. As applied to a hidden-figure item it is quite simple. Basically, the smaller figure is positioned at every possible node of the larger figure, (a node being defined as the intersection of two lines.) The number of lines in the smaller figure that are matched by the larger figure is computed. If, for example, only one side of the smaller figure matches, the count is two; if all sides match, the count is 14. All the smaller figures we used have seven sides; each side counts as two, so a 14 indicates that the smaller figure is embedded in the larger figure. A matrix of counts is generated by this process, in which each element of the matrix corresponds to a count.

Figure 2 shows several items of apparently increasing difficulty. The simplest item yields a matrix of counts, with a 14 surrounded by 2's and 4's. The most difficult item, however, has several 12's surrounding the 14. That is, there were many subfigures surrounding the embedded figure that are very similar to it, and as a result it becomes more difficult to disembed
the smaller figure. When the figure is not there, i.e., for false items, a similar analysis applies. That is, many 12’s in close proximity confuse the viewer into believing that the subfigure is there, when it is not.

---

Insert Figure 2 about here

---

The purpose of this report can be stated as seeking to validate this grammar of the problem. An approach that is consistent with the generative approach is to formulate an item-generation algorithm capable of creating items that have the same underlying matrix of counts but different visual realizations. Eight pairs of items were generated in this fashion by means of a computer program. It is beyond the scope of this paper to discuss the program, but the reader is referred to Ronse & Devijver (1984) for a discussion of a general program that uses a similar but far more general approach to the detection of subfigures. The generation component in our program, although not trivial, is nothing but efficient search.

The item-generation algorithm takes the matrix described above and a small pattern and tries to create a large pattern that matches the matrix. The generation process is simplified by the fact that patterns only contain horizontal, vertical, and 45-degree lines between nodes. The basic idea is to start with a large pattern including all the possible lines and keep removing lines until the matching algorithm produces a matrix that equals the input matrix.

The process starts at the upper left node by calculating all the possible sets of lines that can be removed to make the corresponding matrix value equal the desired value. The program chooses one of these sets,
removing the appropriate lines. This action is repeated for the next and subsequent nodes. One line can affect many matrix values, so the program must make sure that none of these sets contains a line that could make some matrix value go below its desired value. The process continues until the input matrix is matched or the matrix value of some node cannot be made equal to the proper value.

If a node is reached that cannot be made equal to its desired value, the algorithm must backtrack to some previous node and choose some other set of lines to remove. It first backtracks to the node it most recently dealt with that can affect the node it stopped on. If this node cannot be made to match its desired value in another way, the program backtracks further. If no node can be found to backtrack to, the generation process fails.

The Items

Eight items were selected from the Factor Kit (Ekstrom, French, & Harman, 1976) as the generating items. The underlying matrix for each of these was computed. The resulting matrix was then used to generate eight pairs of clones. The eight generating items and the eight pairs of clones appear in Appendix A.

The items were assembled into two forms; the first eight items were common to both forms A and B and consisted of the eight generating items. The last eight consisted of set A of clones for Form A and set B of clones for Form B. The items were positioned in the two forms in such a way that the clones occupied the same position. Form A and Form B were put on an inexpensive graphic microcomputer (a Radio Shack Color Computer) with graphic resolution of 256 x 192. A color monitor (Amdek Color I) was used to display the items. Subjects responded by means of a joystick (Radio Shack No. 26-3012). They were instructed to move the joystick forward if
they thought the item was true and back if they thought the item was false. The instruction for the subject appears in Appendix B. Subjects' reaction time was recorded with 1/60th of a second resolution, and they were informed if they were correct or not after responding to each item.

Subjects

Subjects that participated in the study were high school students from Princeton, New Jersey, and surrounding communities. Sixty students participated, approximately equally distributed between males and females. The data were not edited in any way prior to the analysis presented below. Twenty-nine students took Form A, while thirty-one students took Form B.

Results

We will examine the validity of the proposed grammar by examining the relationship between difficulty estimates for groups A and B on the generating items as well as the clones. To the extent that the grammar is correct the expectation is that the difficulty estimates will not only be linearly related but in addition will fall along a line with slope of 1. Secondly, we will examine an item-by-item analysis of the response-time distribution. Difficulty was estimated by the formula,

\[ \Delta = \log \left( \frac{p}{1-p} \right) \]

Larger values of \( \Delta \) are associated with easier items. Some of the statistical properties of \( \Delta \) have been discussed recently by Holland and Thayer (1985).

As can be seen in Figure 3 the estimated difficulties tend to fall along a diagonal with a slope of 1.0. The correlation between difficulty
estimate was .41. Although not extremely high, the items seem to scatter along the theoretical line with slope of 1.0.

Figure 4 shows the relationship between difficulty estimates to the same two groups responding to a different set of clone items. As can be seen the relationship is strong (correlation of .74) but more importantly the estimates also tend to fall along a diagonal line with slope of 1.0.

If we contrast Figures 3 and 4, we find that there seems to be a significant amount of learning taking place within very few items. The median difficulty of the generating items is approximately 0.5, whereas it is 1.5 for the clones, which were administered subsequent to the generating items. To interpret this effect as learning rather than practice, we should have had a more complex design. Fortunately, these issues are not central to the question of whether we have successfully cloned the items, but we will revisit the issue in the discussion section.

A more stringent assessment of the success of the cloning process goes beyond the comparison of difficulty estimates into an examination of response times. That is, the time it takes to respond would seem to be more informative as to whether or not the same psychological processes are involved in responding to items that are supposed to be psychometric clones. Figure 5 shows the cumulative response time distribution for the eight generating items. By response time we mean the elapsed time until a
positive or negative response was given, regardless of whether it was
correct or incorrect. Each plot in that figure shows the cumulative
distribution for groups A and B together. The expectation is that since
both groups are randomly equivalent the cumulative distribution of response
time will be very close to each other. As can be seen this is true for the
most part. This result is reassuring, but note that in addition to the
close distribution for a given item, the shape of the distributions for the
different items is somewhat different, an effect suggesting that the
response process varies as a function of the item characteristics.

Figure 6 shows the response-time distributions for the two sets of
eight clones. Again, each plot shows the cumulative response-time distri-
bution corresponds to clones rather than to the generating items adminis-
tered to the two groups. The most discrepant item is 8. Items 1 and 2
appear discrepant, but on closer examination it is evident that the
discrepancy is accounted for, the most part, by a couple of the subjects
having taken too long to respond, perhaps the result of some local
distraction. As with the distribution for the generating items, the fact
that there are differences in the shape of the curve across items but not
within clones suggests that essentially the same response processes are
being measured by the clones.
Discussion

A generative approach to psychometric modeling incorporates response modeling, item development, and validation in a coherent and cohesive package. The response modeling and item development become, in effect, a single process once we have written the grammar for the item type in question. To the extent that the grammar is successful we have a means of sampling at random strata of a universe of items such that the psychometric item characteristics of items belonging to a stratum are identical. As it is true of other types of model, the possibility of misspecification exists. Just as a one-parameter logistic model, often used in psychometric work, may not adequately describe responses to a multiple-choice item, it may also occur that the grammar for a particular item type may not adequately clone items. In short, there is no escaping the validation phase. Validation is, in fact, an integral part of the generative approach. First, by basing the grammar on previous research, we are insuring that the items generated using the grammar will be based on that research. In a sense we build in validity. Secondly, the grammar will be tested continually because of the computerized nature of the administration processes assumed by a generative approach. As items are generated, data will be collected on them, and, in the context of computer-administered tests, it should be feasible to maintain a record of the adequacy of the generated items. For example, within an IRT framework, we would assign the same item parameter estimates to items generated from the same generating item (designs for estimating the parameters for generating items are beyond the scope of this paper). Then, in order to see if the assignment is correct we could examine if performance on a generated item fits the parameters of the generating item.
In the results just presented we were not able to obtain guidance from
existing research to help us in the choice of approach to representing the
destined. As a result the findings serve primarily to illustrate the processes
involved in the application of generative psychometrics. The approach we
did take, however, would seem to be compatible with a template-matching
approach. While template matching as a theory of object recognition is not
very tenable (e.g., Pinker, 1984) it does not seem unreasonable as the basic
mechanism for disembedding a smaller figure from a larger one. That is,
performance on both true and false items like the ones used in this investi-
gation is controlled by the position and magnitude of the counts in the
matrix: for true items, the more entries there are approaching 14 in the
immediate neighborhood where a 14 does exist, the longer it would take to
arrive at a decision. Similarly for false items the number and distribution
of counts below 14 would seem to control performance.

The computational flavor of this description is certainly in line with
cognitive psychology but seems to be at odds with Gestalt psychology, which
would claim that perception cannot be understood simply as the sum of the
parts. Some evidence in support of this claim is suggested by the differ-
ence in difficulty between generating items and their corresponding clones.
Although the clones appeared last in the test it is not likely that their
lower difficulty is just a position effect. An alternative explanation is
suggested by an examination of the generating and their clones (see Appendix
A) which shows that a global feature of the generating item that is not
preserved by the generation algorithm is symmetry. Symmetry is known to
play an important role in the recall, recognition and discrimination of
figures (Attneave, 1955; Adams, Fitts, Rappaport, & Weinstein, 1954; Soltz &
Wertheimer, 1959; Chipman 1977; Royer, 1981). It is thus possible that the
In the results just presented we were not able to obtain guidance from existing research to help us in the choice of approach to representing the item. As a result the findings serve primarily to illustrate the processes involved in the application of generative psychometrics. The approach we did take, however, would seem to be compatible with a template-matching approach. While template matching as a theory of object recognition is not very tenable (e.g., Pinker, 1984) it does not seem unreasonable as the basic mechanism for disembedding a smaller figure from a larger one. That is, performance on both true and false items like the ones used in this investigation is controlled by the position and magnitude of the counts in the matrix: for true items, the more entries there are approaching 14 in the immediate neighborhood where a 14 does exist, the longer it would take to arrive at a decision. Similarly for false items the number and distribution of counts below 14 would seem to control performance.

The computational flavor of this description is certainly in line with cognitive psychology but seems to be at odds with Gestalt psychology, which would claim that perception cannot be understood simply as the sum of the parts. Some evidence in support of this claim is suggested by the difference in difficulty between generating items and their corresponding clones. Although the clones appeared last in the test it is not likely that their lower difficulty is just a position effect. An alternative explanation is suggested by an examination of the generating and their clones (see Appendix A) which shows that a global feature of the generating item that is not preserved by the generation algorithm is symmetry. Symmetry is known to play an important role in the recall, recognition and discrimination of figures (Attneave, 1955; Adams, Fitts, Rappaport, & Weinstein, 1954; Soltz & Wertheimer, 1959; Chipman 1977; Royer, 1981). It is thus possible that the
In the results just presented we were not able to obtain guidance from existing research to help us in the choice of approach to representing the item. As a result the findings serve primarily to illustrate the processes involved in the application of generative psychometrics. The approach we did take, however, would seem to be compatible with a template-matching approach. While template matching as a theory of object recognition is not very tenable (e.g., Pinker, 1984) it does not seem unreasonable as the basic mechanism for disembedding a smaller figure from a larger one. That is, performance on both true and false items like the ones used in this investigation is controlled by the position and magnitude of the counts in the matrix: for true items, the more entries there are approaching 14 in the immediate neighborhood where a 14 does exist, the longer it would take to arrive at a decision. Similarly for false items the number and distribution of counts below 14 would seem to control performance.

The computational flavor of this description is certainly in line with cognitive psychology but seems to be at odds with Gestalt psychology, which would claim that perception cannot be understood simply as the sum of the parts. Some evidence in support of this claim is suggested by the difference in difficulty between generating items and their corresponding clones. Although the clones appeared last in the test it is not likely that their lower difficulty is just a position effect. An alternative explanation is suggested by an examination of the generating and their clones (see Appendix A) which shows that a global feature of the generating item that is not preserved by the generation algorithm is symmetry. Symmetry is known to play an important role in the recall, recognition and discrimination of figures (Attneave, 1955; Adams, Fitts, Rappaport, & Weinstein, 1954; Soltz & Wertheimer, 1959; Chipman 1977; Royer, 1981). It is thus possible that the


Figure 1

Sample Hidden Figure Items

True Item

False Item
Figure 2

Hidden Figure Item of Increasing Complexity

Matrix

0 2 2 2 2
2 0 0 2 2
4 4 4 4 4
2 2 2 14 2 4
2 2 2 2 2 4
0 0 2 2 0 0
2 4 4 4 4 2

Matrix

0 2 2 2 2 2 0
2 0 2 4 4 4 2
6 6 14 8 8 8 4
4 2 6 14 4 6 4
4 2 6 6 6 6 4
2 2 6 6 4 2 2
2 4 4 4 4 2 0

Matrix

0 2 2 2 2 2 0
2 4 4 4 4 4 0
6 10 12 12 12 12 8 4
4 8 6 16 6 8 0
6 10 12 12 12 10 4
2 4 4 4 4 2 0
2 4 4 4 4 2 0
Figure 3

Relationship Between Difficulty Estimates of Generating Based on Groups A and B
Figure 4

Relationship Between Difficulty Estimates for Pairs of Clones from a Common Generating Item Administered to Random Groups A and B
Cumulative response time distributions for the eight generating items administered to two random groups A and B. The relative item position is indicated below the figure label.
Figure 6
Cumulative response time distributions for the eight pairs of clones administered to random groups A and B. The relative item position is indicated below the figure label.
APPENDIX A

Eight Generating Items with Clones a and b
Generating Item

Clone a

Clone b
HF 8
Generating Item

Clone a

Clone b
HF 9
Generating Item

Clone a

Clone b
HF 10
Generating Item

Clone a

Clone b
HF 12
Generating Item

Clone a

Clone b
HF 14
Generating Item

Clone a

Clone b
APPENDIX B

Instructions for Hidden-Figure Items
Instruction for Hidden Figure Items

In this exercise your task will be to decide whether or not a smaller figure is part of a larger figure. It is important to be FAST and CORRECT. To see an example where the figure on the right IS PART of the figure on the left press the red button.

(A true item appears on the screen with a blinking hidden figure.)

To see an example where the figure on the right IS NOT PART of the figure on the left press the red button.

A false item appears on the screen.

You are now ready to respond to some practice trials. You must respond QUICKLY and CORRECTLY. However, you can pace yourself because with the red button you control when to see the next trial. The time you take between trials is not counted. There are four practice trials.
Respond CORRECTLY and QUICKLY.
PUSH joystick FORWARD if
the right figure IS part of the left figure.
FULL joystick BACKWARD if
the right figure IS NOT part of the left figure.
Press the red button when you are ready for the
next trial.

(Four practice items are presented.)

You are now ready for the real test. Remember:
PUSH joystick FORWARD if
the right figure IS part of the left figure.
FULL joystick BACKWARD if
the right figure IS NOT part of the left figure.
Press the red button when you are ready for the
next trial.
Personnel Analysis Division,
AF/MPXA
5C360, The Pentagon
Washington, DC 20330

Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

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

Dr. Phipps Arabie
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Technical Director, ARI
5001 Eisenhower Avenue
Alexandria, VA 22333

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

Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

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

Dr. Arthur S. Blaiwes
Code N711
Naval Training Systems Center
Orlando, FL 32813

Dr. Bruce Bloxom
Administrative Sciences
Code 54B1
Navy Postgraduate School
Monterey, CA 93943-5100

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

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

Dr. Robert Breaux
Code N-095R
Naval Training Systems Center
Orlando, FL 32813

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

Dr. Patricia A. Butler
OERI
555 New Jersey Ave., NW
Washington, DC 20208

Mr. James W. Carey
Commandant (G-PTE)
U.S. Coast Guard
2100 Second Street, S.W.
Washington, DC 20593

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

Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514
Dr. Gerhard Fischer  
Liebigasse 5/3  
A 1010 Vienna  
AUSTRIA

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. Carl H. Frederiksen  
McGill University  
3700 McTavish Street  
Montreal, Quebec H3A 1Y2  
CANADA

Dr. Robert D. Gibbons  
University of Illinois-Chicago  
P.O. Box 6998  
Chicago, IL 60680

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

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

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

Dr. Ronald K. Hambleton  
Prof. f Education & Psychology  
University of Massachusetts  
at Amherst  
Hills House  
Amherst, MA 01003

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

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

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

Dr. Paul Horst  
677 G Street, #184  
Chula Vista, CA 90010

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

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

Dr. Steven Hunka  
Department of Education  
University of Alberta  
Edmonton, Alberta  
CANADA

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

Dr. Robert Jannarone  
Department of Psychology  
University of South Carolina  
Columbia, SC 29208

Dr. Douglas A. Jones  
P.O. Box 6640  
Lawrenceville  
NJ. 08648
Dr. William E. Nordbrock  
FMC-ADCO Box 25  
APO, NY 09710

Dr. Melvin R. Novick  
356 Lindquist Center  
for Measurement  
University of Iowa  
Iowa City, IA 52242

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

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

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

Dr. James Olson  
WICAT, Inc.  
1875 South State Street  
Orem, UT 84057

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

Special Assistant for Marine Corps Matters,  
ONR Code OOMC  
800 N. Quincy St.  
Arlington, VA 22217-5000

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

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

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

Dr. Roger Pennell  
Air Force Human Resources Laboratory  
Lowry AFB, CO 80230

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

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

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

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

Dr. Fumiko Samejima  
Department of Psychology  
University of Tennessee  
Knoxville, TN 37916

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

Dr. Robert Sasmor  
HQDA DAMA-ARL  
Pentagon, Room 3E516  
Washington, DC 20310-0631  
USA

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

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