A RAPID ITEM-SEARCH PROCEDURE FOR BAYESIAN ADAPTIVE TESTING

C. David Vale
and
David J. Weiss

RESEARCH REPORT 77-4
MAY 1977

PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

Prepared under contract No. N00014-76-C-0243, NR150-382
with the Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 77-4	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Rapid Item-Search Procedure for Bayesian Adaptive Testing		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) C. David Vale and David J. Weiss		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0243
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, Minnesota 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.:61153N PROJ.:RR042-04 T.A.:RR042-04-01 W.U.:NR150-382
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE May 1977
		13. NUMBER OF PAGES 14
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

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

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

testing	sequential testing	programmed testing
ability testing	branched testing	response-contingent testing
computerized testing	individualized testing	automated testing
adaptive testing	tailored testing	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An alternative item-selection procedure for use with Owen's Bayesian adaptive testing strategy is proposed. This procedure is, by design, faster than Owen's original procedure because it searches only part (as compared with all) of the total item pool. Item selections are, however, identical for both methods. After a conceptual development of the rapid-search procedure, the supporting mathematics are presented. In a simulated comparison with three item pools, the rapid-search procedure required as little as one-tenth the computer time as Owen's technique.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CONTENTS

Technical editing by Terryl Graham

A Rapid Item-Search Procedure for Bayesian Adaptive Testing

In recent years, a number of strategies for administering adaptive tests have been developed. Among the more elegant of these is the strategy developed by Owen (1969, 1975). This strategy is based on a statistical model developed from Bayes' theorem (Phillips, 1973) and modern test theory (Lord & Novick, 1968). At the beginning of test administration under this strategy, an initial estimate of the testee's ability is needed. This is operationalized as a mean (reflecting the test administrator's estimate of a testee's ability level) and a variance (reflecting the confidence the administrator places on the estimate) of a normal-shaped prior ability distribution. In the absence of any prior information about the testee, the prior distribution may be simply the distribution of ability in the population from which the testee was sampled.

During the course of testing, the goal of Owen's strategy is to refine the initial ability estimate. Given the prior distribution, this goal is approached by choosing as the first item to administer the item in a pool of items that is expected to best refine the ability estimate. Having administered this item, a new ability estimate is calculated from the prior ability distribution and the item response. This posterior distribution then becomes a new prior distribution, and the process of item selection, administration, and scoring is repeated. The process continues until either a certain degree of refinement is attained or a pre-specified number of items have been administered.

Because of the complicated calculations required as each item is administered, Owen's Bayesian adaptive testing strategy must be administered by computer. However, the amount of calculation required between items is still great enough that substantial time delays may occur between items. This is due partially to the calculations required to refine the ability estimate after each item is administered. But to a much greater extent, it is due to the inefficient procedure suggested by Owen for finding the most appropriate item to administer. Since Owen's item-search procedure works best with large item pools (Urry, 1971), and because the time it requires increases with increasing item-pool size, the search time required to select the appropriate item at any stage will be large for properly constituted item pools. Although delays between item administrations will have no direct effect on the psychometric properties of the procedure, they might well introduce undesirable psychological effects on test scores (e.g., Betz & Weiss, 1976*a*, 1976*b*).

This paper reviews the conceptual and mathematical bases of Owen's item-search procedure and proposes a more efficient and much faster technique that is particularly useful with large item pools.

Owen's Original Procedure

At each stage of the testing process, Owen's strategy seeks to administer that item which minimizes the expected variance of the posterior ability distribution. This may be accomplished by minimizing what Owen refers to as the *beta* (β) *function*.

Where i indexes an item, let:

$a_i \equiv$ normal ogive discrimination index of item i,

$b_i \equiv$ normal ogive difficulty index of item i,

$c_i \equiv$ probability of a correct response due to random guessing on item i,

$\mu_o \equiv$ mean of the hypothesized normal prior ability distribution,

$\sigma_o^2 \equiv$ variance of the hypothesized normal prior ability distribution,

$$D=(b_i-\mu_o)\Big/\sqrt{2(a_i^{-2}+\sigma_o^2)} \qquad [1]$$

$$K^{-1} = \tfrac{1}{2}(1-ERFN(D)), \qquad [2]$$

and

$$ERFN(x) = \frac{2}{\sqrt{\pi}}\int_o^x e^{-t^2}\,dt \qquad [3]$$

Then:

$$\beta_i = (1-c_i)^{-1}\,(1+\sigma_o^{-2}a_i^{-2})\,(1-K^{-1})\,[c_i+(1-c_i)K^{-1}]e^{2D^2}. \qquad [4]$$

β_i is a function of five variables: a_i, b_i, c_i, μ_o, and σ_o^2. When searching for an item, the prior distribution and, thus, μ_o and σ_o^2, are constant. For convenience, c_i is also usually assumed to be constant. Therefore, when searching for an item to administer, β_i is a function of only a_i and b_i.

Figure 1 is a plot of the values of the beta function for 313 items from a real item pool plotted as a function of a and b with μ_o, σ_o^2 and c respectively fixed at 0, 1, and .2. Given a finite pool of items such as this, Owen suggested calculating the beta value for each item and choosing the item for which that value was a minimum. This amounts to (symbolically) generating a plot like that shown in Figure 1, and choosing the item corresponding to the lowest dot.

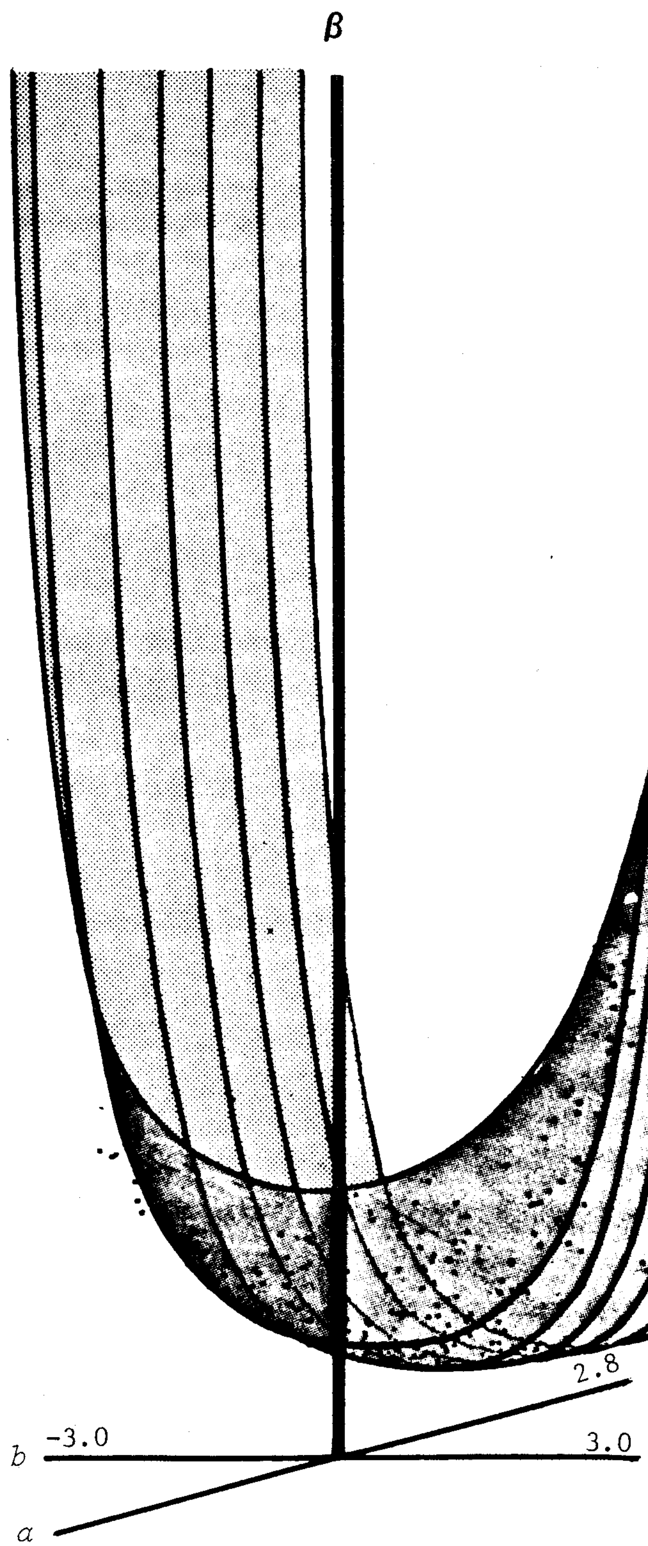
With a pool of 500 items, Owen's search procedure may require over five seconds of computer time for each item selected on a relatively sophisticated minicomputer. This is equivalent to over five minutes of computer time just to select items for a 60-item test. In a simulation study such as that reported by McBride and Weiss (1976), selecting items for the 15,000 simulated subjects needed to calculate one information curve would take over two weeks of computer time if a real item pool were used. Obviously, some refinement in the search procedure would be welcome, for use in both live-testing studies and computer simulation studies with real item pools.

A More Efficient Search Procedure

Conceptualization

In Figure 1, it may be noted that the low dots (i.e., items) appear in one area of the plot and that the dots get higher as a function of the distance from

Figure 1
Beta Values of 313 Test Items with $\mu_o=0$, $\sigma_o^2=1.0$, and $c=.2$



that area. If Figure 1 is viewed as a continuous plot of the beta function, for every value of a, there is one value, M, of b for which β is minimum. β appears to be a monotonic increasing function of $|b-M|$, and a monotonic decreasing function of a. These observations can be combined to create a more efficient item-search strategy.

The best item for minimizing beta will be a highly discriminating item with difficulty of $b=M$. Therefore, an efficient item search should begin at the point where $b=M$ and a is at the upper bound of item discrimination in the pool. The search could then proceed by first evaluating items close to that point and then working outward, while keeping track of the beta value of the best item yet found. The search should end when no item in the area of the plot yet unsearched could possibly have a lower beta value than the currently best item.

The point at which no possibly better items remain can be determined, conceptually at least, by plotting an iso-beta contour (a curve described by the intersection of a plane parallel to the a-b plane with the beta surface, like the curve shown in Figure 2) through the currently best item. All points within the curve have lower beta values than any points outside the curve. Therefore, when all the area inside the curve has been searched, no better items will be found.

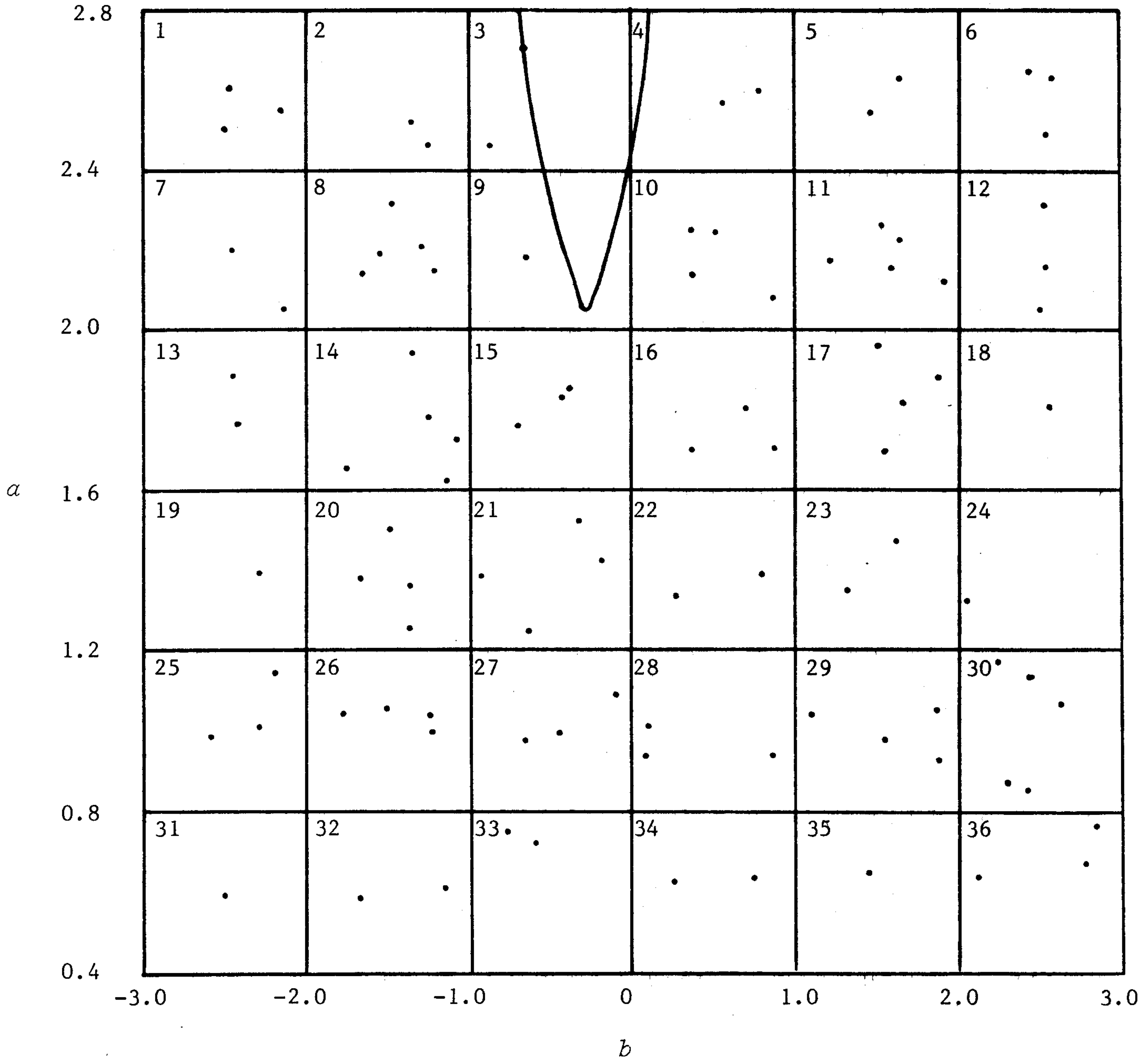
Unfortunately, a digital computer is not equipped to handle this conceptual graphic search very well, so a discrete approximation must be implemented. This is accomplished by blocking the $a \times b$ item-pool plot into rectangles and searching the rectangles one at a time. Figure 2 shows an item pool plot so divided with each block numbered for ease of reference.

Example

The search procedure was implemented in the blocked item pool shown in Figure 2. With μ_o, σ_o^2, and c defined as before, when M was evaluated at a=2.8 (the a-value of the most discriminating item in this pool), M=-.274; thus, the search began in block 3, which contained two items, the better item having a beta value of .440. The conceptual iso-beta contour is plotted through this item in Figure 2. The boundary values of beta at b=-1.0 with a=2.8 and 2.4 were evaluated, and it was determined that all lower blocks in row 1 (blocks 1 and 2) fell outside the iso-beta contour and thus were not searched. The upper boundaries of block 3 were then evaluated and block 4 was searched. No better items were found in block 4. The upper boundaries of block 4 were evaluated, and it was determined that no higher blocks in row 1 could contain better items, so they were not searched.

Next, a new value of M, with a fixed at 2.4, was calculated to be -.280, and block 9 was searched but no better items were found. The upper boundary of block 8, at b=-1.0, and the lower boundary of block 10, at b=0.0, were evaluated. These boundaries were both outside the iso-beta contour and therefore, no more blocks in this row were searched. A new value of M was calculated at a=2.0 and the beta at that point was found to be .453, a value higher than that of the currently best item. Since this was the minimum value of beta that could be obtained with items of a=2.0 or less, the remainder of the item pool was not searched. In all, three of the 36 blocks were searched.

Figure 2
A Blocked Item Pool



Mathematics Necessary for the Procedure

M is found by setting the first partial derivative of β with respect to b equal to zero for the given value of a.

Let:

$$F=(1-c)^{-1}(1+\sigma_o^{-2}a^{-2}) \qquad [5]$$

$$H=[c+(1-c)K^{-1}]e^{2D^2}. \qquad [6]$$

Then:

$$\beta=F(1-K^{-1})H. \qquad [7]$$

For fixed a, c, and σ_o^2, F is a constant. Therefore:

$$\frac{\partial\beta}{\partial b} = F[(1-K^{-1})\frac{\partial H}{\partial b} + H\frac{\partial(1-K^{-1})}{\partial b}] \qquad [8]$$

$$\frac{\partial(K^{-1})}{\partial b} = -\frac{e^{-D^2}}{\sqrt{\pi}}[2(a^{-2}+\sigma_o^2)]^{-\frac{1}{2}} \qquad [9]$$

$$\frac{\partial(1-K^{-1})}{\partial b} = -\frac{\partial(K^{-1})}{\partial b} = \frac{e^{-D^2}}{\sqrt{\pi}}[2(a^{-2}+\sigma_o^2)]^{-\frac{1}{2}} \qquad [10]$$

$$\frac{\partial H}{\partial b} = \frac{1}{\sqrt{2(a^{-2}+\sigma_o^2)}}\left[H\cdot 4D + \frac{e^{D^2}}{\sqrt{\pi}}(c-1)\right]. \qquad [11]$$

Expanding and rearranging:

$$\frac{\partial\beta}{\partial b} = F\cdot(1-K^{-1})\frac{e^{D^2}}{\sqrt{2(a^{-2}+\sigma_o^2)}}\left([c+(1-c)K^{-1}]\cdot\right.$$

$$\left.\left(4De^{D^2} + \frac{1}{\sqrt{\pi}(1-K^{-1})}\right) + \frac{1}{\sqrt{\pi}}(c-1)\right), \qquad [12]$$

which is equal to zero if and only if

$$Q=\left([c+(1-c)K^{-1}]\left(4De^{D^2} + \frac{1}{\sqrt{\pi}(1-K^{-1})}\right) + \frac{1}{\sqrt{\pi}}(c-1)\right) \qquad [13]$$

is equal to zero.

The root of Q at which $\frac{\partial\beta}{\partial b} = 0$ and β is a minimum can easily be found using the Newton-Raphson iteration. The derivation of Q needed in the procedure is given below.

Let: $S=(c + (1-c)K^{-1})$ [14]

$$T=\left(4De^{D^2} + \frac{1}{\sqrt{\pi}(1-K^{-1})}\right) \qquad [15]$$

$$Q=ST + \frac{1}{\sqrt{\pi}}(c-1) \qquad [16]$$

$$\frac{\partial Q}{\partial b} = S\frac{\partial T}{\partial b} + T\frac{\partial S}{\partial b} \qquad [17]$$

$$\frac{\partial S}{\partial b} = (c-1)\Big/\left(e^{D^2}\sqrt{2\pi(a^{-2}+\sigma_o^2)}\right) \qquad [18]$$

$$\frac{\partial T}{\partial b} = \frac{e^{D^2}}{\sqrt{2(a^{-2}+\sigma_o^2)}} \cdot \left(4+8D^2 - \frac{e^{-2D^2}}{\pi(1-K^{-1})^2}\right) \qquad [19]$$

Block Size

Using (μ_o-c) as the initial value of b, the Newton-Raphson procedure typically converges to $\Delta b<.01$ in two cycles and to $\Delta b<.0001$ in four. The precision needed is dependent on the size of blocks used. With block widths of $0.5b$ and $0.3a$, no deficit in performance was noted (as would be evidenced by the rapid search prodecure choosing an item different from the one chosen by the full search procedure) when a convergence criterion as crude as $\Delta b<.1$ was used. The danger in using a crude estimate of M is that the search may stop a row too soon and miss a good item. If a few misses could be tolerated, some time would be saved by accepting as the minimum beta for a level of a that value of β obtained when evaluated at $b=\mu_o-c$. For research purposes, this may not be tolerable, however, and the value of beta at $b=M$ must be determined.

The equations necessary to determine the optimal size and spacing of the blocks in the $a \times b$ grid have not been developed. Conceptually, it seems that--given a pool of items and some assumptions about the distribution of ability in the testee population--there should be an optimal size for each block to minimize the required search time. But in the absence of the mathematically optimal solution, there are two relevant considerations. First, each block will require additional computer memory. Furthermore, the procedure requires an amount of computer time slightly greater than that required to evaluate one item in order to determine whether a block could conceivably contain a better item.

Timing Comparisons in Three Item Pools

For timing comparisons reported below, grids of two levels of resolution were used. For a small item pool containing 200 items, a 48-block grid (six levels of a and eight levels of b) was used. For two larger item pools containing 313 and 580 items, a 96-block grid (eight levels of a and twelve levels of b) was used. These sizes were chosen somewhat arbitrarily. An optimal grid size should produce comparisons more favorable to the partial search technique.

Table 1 shows timing statistics for both Owen's full search technique and the rapid search technique in three item pools. The basic item pool from which these items (actually item statistics) were drawn was a real pool of 569 items (McBride & Weiss, 1974). The 313-item pool consisted of those items with b-values between ±3.0 and a-values between 0.4 and 2.8. To evaluate the relative efficiency of the two search techniques for a current project using a 200-item pool, 200 items were randomly sampled from the 313. The 580-item pool contains the item statistics obtained from the 313-item pool and 267 additional sets of item statistics obtained from an earlier calibration of the same items. These three pools are shown in blocked form in the Appendix.

Table 1
Timing Statistics for Two Search Procedures

No. Items	Grid Size		Average Search Time per Test		Rapid as Percent of Full	Time per Evaluation (Full Search)	Item Equivalent
	a	b	Full	Rapid			
200	0.4	0.75	3.195	1.071	33.526	574*	76.932
313	0.3	0.50	5.118	.976	19.080	571*	71.404
580	0.3	0.50	9.705	1.020	10.512	572*	73.947

*Time per item in microseconds

Columns three and four of Table 1 show time in seconds required by a Control Data Corporation 6400 computer, using the two procedures, to select 30 items. These items were selected during a computer simulation of the Bayesian test (see McBride and Weiss, 1976 for details of the simulation procedure). For each time value shown in Table 1, 100 testees were simulated, sampling ability levels from a normal distribution with mean of zero and standard deviation of 1.0. Table 1 shows the average search time required by Owen's full search procedure and the rapid search procedure to administer a thirty-item test to each simulated testee. Column six in Table 1 shows the percentage of time taken by the rapid search procedure relative to the full search procedure. With a relatively small pool (200 items) to search and a rough grid (6×8), the rapid search technique was three times faster than the full search procedure. With a larger pool (580 items) and a finer grid (8×12), the rapid search was almost ten times as fast.

Another way of comparing the relative efficiency of the two procedures is by comparing the relative sizes of pools that can be searched in a given time. Let:

$I \equiv$ the number of items to be administered
$J \equiv$ the number of items in the pool
$E \equiv$ the number of item evaluations performed
$T \equiv$ the time spent in selecting the I items
$t \equiv$ the time required to evaluate one item

Since at each stage of the test, one item is eliminated and thus not evaluated in further searches:

$$E=J+(J-1)+(J-2)+ \cdots +(J-(I-1))$$

$$=IJ- \sum_{i=1}^{(I-1)} i$$

$$=IJ- \frac{I(I-1)}{2} \qquad [20]$$

and

$$t=T/E \qquad [21]$$

The time to evaluate one item in the full search procedures, t, should be constant across item pools of varying sizes, and, as shown in column seven of Table 1, is nearly constant with a median of 572 microseconds.

Substituting and rearranging:

$$J=T/(tI)+ \frac{I-1}{2} \qquad [22]$$

Using .000572 for t, 3 for I, and the time values of column five in Table 1 (i.e., the time taken by the rapid search procedure) for T results in the values shown in column eight, the size of the item pool that could have been searched using the full search procedure in the amount of time taken by the rapid search to effectively search the entire pool. Although the values are crude because of the non-optimal block sizes used, it appears from column eight that by using the rapid search procedure, an item pool of up to about 600 items can be searched in the amount of time required by the full search procedure to search a pool of about 80 items. Since 80 items are probably too few to allow the Bayesian procedure to perform well with a 30-item test, this means that if time is available to administer a Bayesian test, then a relatively large item pool can be used without increasing computer time, if the rapid search procedure is implemented. Since the fidelity of Owen's procedure is a function of the number of items available from which to choose, given a fixed testing time, the rapid search procedure will result in higher test validites if a large item pool is available.

Conclusions

Data presented suggest that the proposed rapid search procedure can accomplish the task performed by Owen's full search procedure as well as the full search procedure in as little as ten percent of the time when used with item pools of typical size. There are two practical needs for this time saving: In live testing, when four subjects are being tested by a minicomputer, a five-second item search time can result in a presentation latency of up to 20 seconds when all testees respond at once or close to each other. This may be sufficient time for a testee to get bored and lose interest in the test. In computer simulations of testing, two weeks is too long to wait for one information curve. Three days (a weekend) for two is tolerable.

Three areas of future research related to Bayesian item pool search techniques are open. First, relative to the rapid search technique, several relationships between a, b, and β were assumed but not proved. Although the relationships seem appropriate, rigorous proofs would be welcome. Second, a method for determination of the optimal grid size as a function of the item pool and an assumed prior ability distribution was not developed. This could further speed up the rapid search procedure. Finally, the degradation in performance of the Bayesian testing strategy using a simpler item evaluation technique should be evaluated. It is possible that simply choosing items of the appropriate difficulty would provide nearly as efficient a test with much less computer time being required.

REFERENCES

Betz, N. E., & Weiss, D. J. Effects of immediate knowledge of results and adaptive testing on ability test performance. (Research Rep. 76-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1976. (AD A027147) (*a*)

Betz, N. E., & Weiss, D. J. Psychological effects of immediate knowledge of results and adaptive ability testing. (Research Rep. 76-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1976. (AD A027170) (*b*)

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.

McBride, J. R., & Weiss, D. J. A word knowledge item pool for adaptive ability measurement. (Research Rep. 74-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1974. (AD 781894)

McBride, J. R., & Weiss, D. J. A word knowledge item pool for adaptive ability measurement. (Research Rep. 74-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1974. (AD 781894)

McBride, J. R., & Weiss, D. J. Some properties of a Bayesian adaptive ability testing strategy. (Research Rep. 76-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1976. (AD A022964)

Owen, R. J. A Bayesian approach to tailored testing. (Research Bulletin RB-69-92). Princeton, NJ: Educational Testing Service, 1969.

Owen, R. J. A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. Journal of the American Statistical Association, 1975, 70(350), 351-356.

Phillips, L. D. Bayesian statistics for social scientists. New York: Crowell, 1973.

Urry, V. W. Individualized testing by Bayesian estimation. (Research Bulletin 0171-177). Seattle: University of Washington, Bureau of Testing, 1971.

Figure A-1
A 48-Block Grid Containing the 200-Item Pool

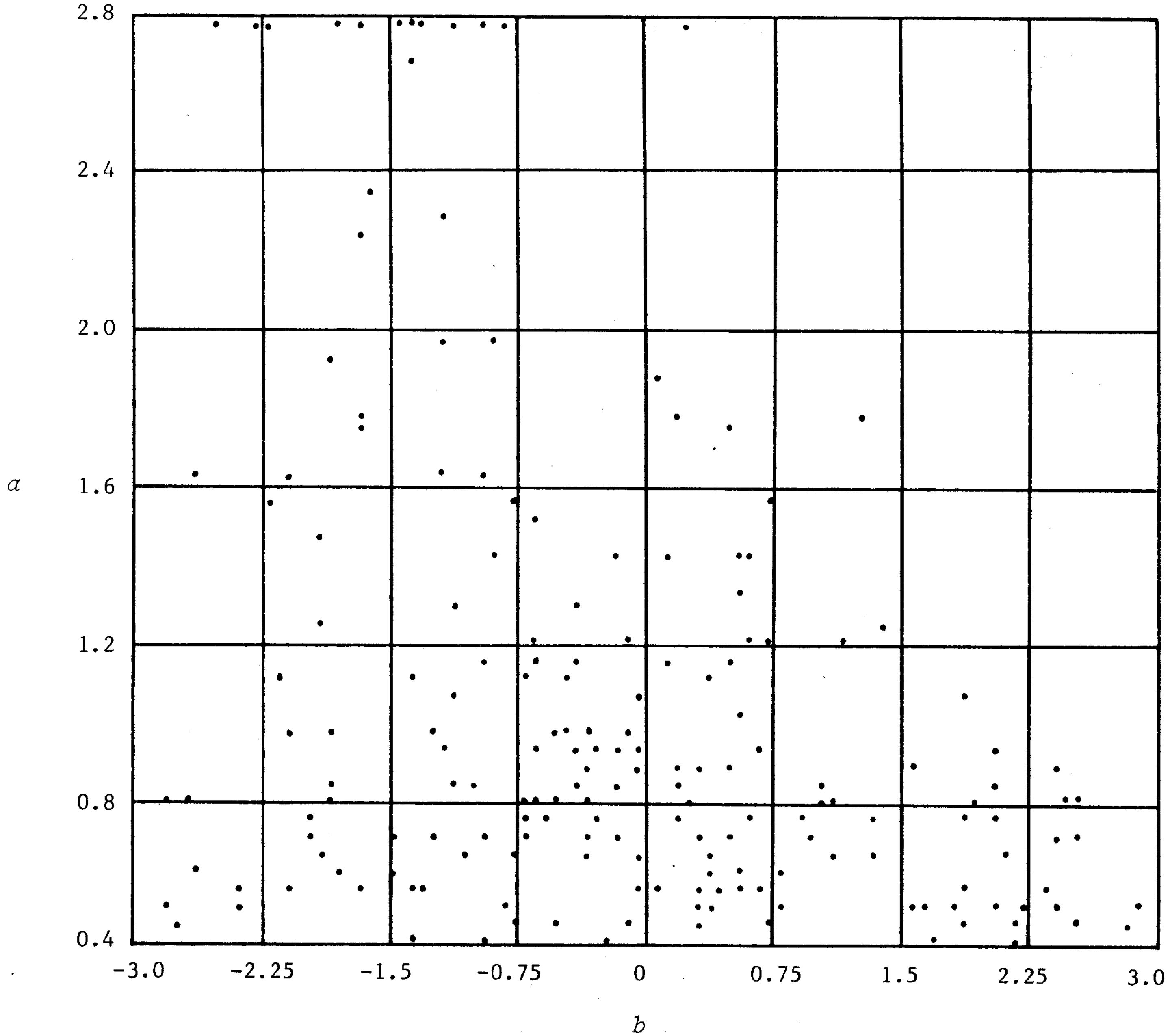


Figure A-2
A 96-Block Grid Containing the 313-Item Pool

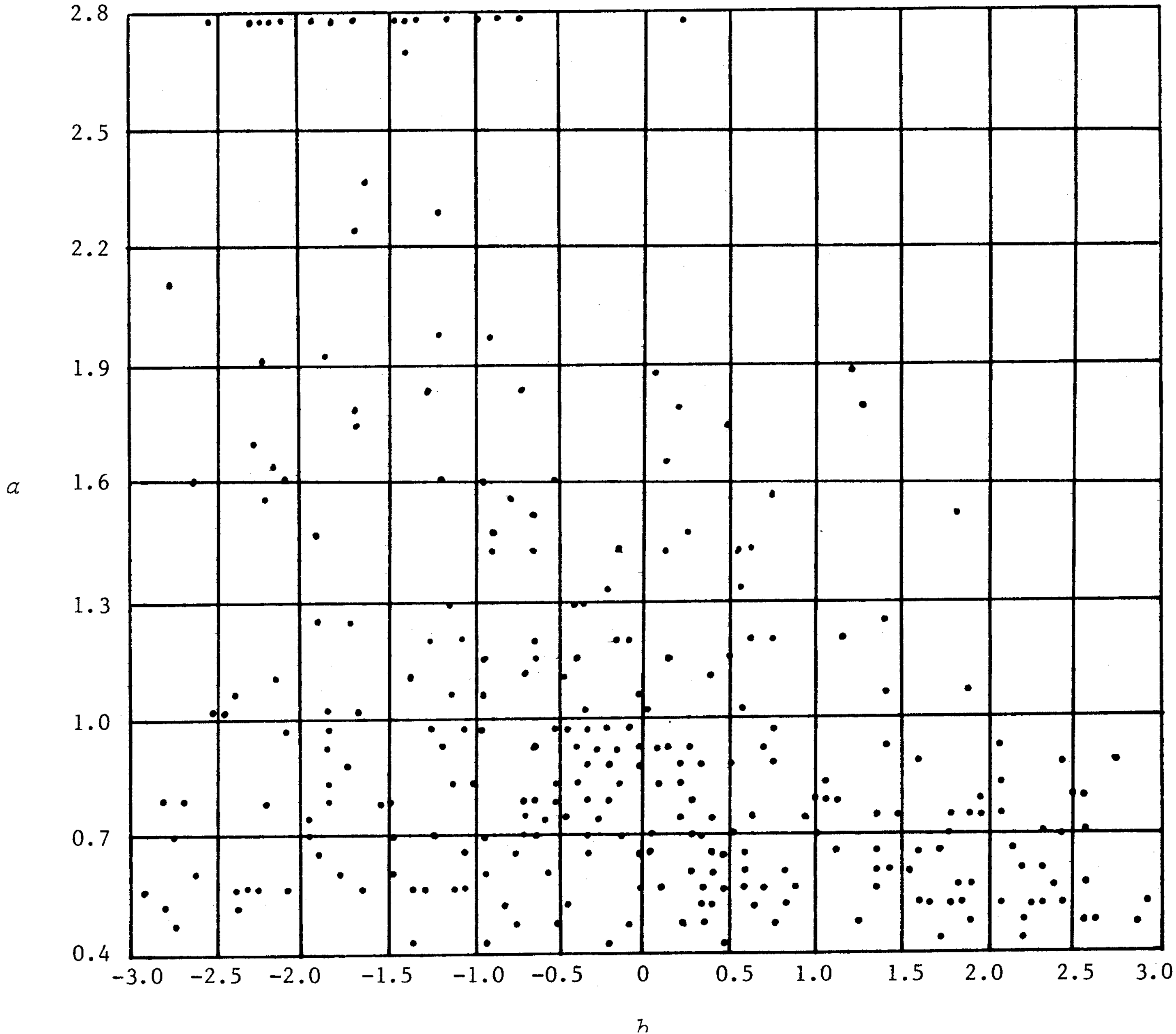


Figure A-3
A 96-Block Grid Containing the 580-Item Pool