

ASSESSING RETARDED BEHAVIOR:
ACCURACY AND TEMPORAL DIMENSIONS

Beatrice H. Barrett

Working draft, prepared July 1978, adapted from:

Annual Reports of the Behavior Prosthesis Department, 1 July 1970 - 30 June 1971, 1 July 1972 - 30 June 1973, 1 July 1973 - 30 June 1974, 1 July 1974 - 30 June 1975. Belmont, Mass.: Walter E. Fernald State School, 1971, 1973, 1974, 1975.

Acquisition subgroups: A step toward functional assessment of learning potential among the severely and profoundly retarded. Paper presented at annual meeting of American Academy on Mental Retardation, Atlanta, May 1973.

The course of behavioral acquisition: Implications for assessment and training of the severely retarded. Paper presented at annual meeting of American Academy on Mental Retardation, Portland, Ore., May 1975.

In developing our methodology for assessing behavioral retardation, we considered a fundamental task to be that of amplifying individual behavioral differences. This seemed especially important with a population that had become homogenized as a result of insensitive assessment methods, assumed upper limits of behavioral competence, and uniform societal rejection.

Over the years our studies revealed reliably different individual behavior patterns ranging from "normal" (as shown by presumably normal children and adults) to severely defective or deficient. We found this broad range not only among people classified as moderate to borderline in retardation (Barrett, 1965, 1969) but also among those classified as severely and profoundly retarded. These individual differences emerged when each participant was given enough exposure to a constant program of differential reinforcement to reach his or her own maximally efficient performance.

Since efficiency of habilitative endeavors is generally thought to be enhanced by supposedly homogeneous groupings, another task of assessment methodology is to locate among individualized behavior patterns the commonalities that describe functional behavior subgroups. In our automatically programmed environment that differentially reinforced both response differentiation and stimulus discrimination, six distinct acquisition "states" eventually emerged from the behavior of our participants. Their psychometric classifications (moderate to borderline versus severely and profoundly retarded) were not related to their asymptotic performance accuracy, but even when performing at maximum accuracy, the severely and profoundly retarded participants responded at significantly lower rates than the less retarded participants. In addition, they took significantly longer to show their maximally efficient levels.

In this report, we describe the standard contingencies under which the six acquisition subgroups emerged. We compare the two psychometric groups with respect to their distribution among the acquisition subgroups, the time they required to show their best performance, and the rates at which they performed most accurately. The relationships of accuracy, acquisition time, and overall rate to age, sex, age at admission, and duration of institutionalization are also analyzed. We then compare the two psychometric groups with respect to successive stages in the emergence of their most accurate performance and with respect to the rates at which they engaged in reinforced behavior during the course of acquisition. More detailed information about the course of acquisition should help increase the prognostic efficiency of assessment procedures and contribute to formulation of more "normalizing" training objectives.

Participants

When they entered our studies, the 64 participants, 43 boys and 21 girls, ranged in age from 3.8 years to 19.8 years (mean: 11.4 years). They had been at Fernald for periods ranging from one day to 14.1 years (mean: 6.1 years). Their ages at admission ranged from 6 days to 18.2 years (mean: 5.4 years). According to the AAMD system (Heber, 1959), 44% were classified as Levels III-V (moderate to borderline) and 56% as Levels I-II (severely or profoundly retarded). Institution psychometrists considered 31% of the group to be "untestable."

The only residents deliberately excluded from our studies were those who, because of physical handicaps, could not operate our differentiation/discrimination assessment apparatus. Children with behavior problems (including indiscriminate defecation and urination) were not denied participation.

Behavior Analytic Environment

Studies were conducted in a sound-attenuated cubicle, where a participant could work independently with no one else present. A wall-mounted console provided two plungers (operable with 300 grams of force through a 1-cm excursion) and an aperture into which objects were automatically dispensed as potential reinforcement for appropriate responding. Two display panels, one above each plunger, were alternately illuminated for one-minute periods throughout one-hour assessment sessions. A candy or a token (later exchangeable for soft drinks) could be earned by every tenth operation of the left plunger only during periods when the left light was on. When the light on the right side was on, nothing could be earned by operating either of the plungers. No penalties were programmed for nonreinforced responses, so a participant could obtain reinforcement in a variety of ways. There was no "pretraining." Introduction to the apparatus consisted of demonstrating plunger operation when the power was off. The conditions remained in effect throughout successive sessions, held from one to three times a week, until the participant's record showed four consecutive sessions at asymptotic performance accuracy, indicating that he or she had learned as much as possible under the existing contingencies.

With two plungers and two lights, four potential stimulus-response combinations were available. Acquisition was shown by an increase in the rate of the one reinforced combination relative to the rates of each of those that were not reinforced (Table 1). These rate relationships were quantified by a rate index, defined as the difference between two rates divided by their sum. Thus, the difference between the reinforced rate and each of the unreinforced rates yielded three rate indices for each participant for each session.

Subgroups Defined by Acquisition of Differentiation and Discrimination

Figure 1 shows the ranges of rate indices that describe the six subgroups that emerged from repeated studies. These patterns of behavior represent criteria of acquisition, specific deficit, and deficiency that were determined empirically and calibrated functionally by the behavior of the target population rather than by a nonretarded normative population (Barrett, 1965). "Differentiation problem," "discrimination problem," "overgeneralization," and "full acquisition" describe four states of process acquisition. Sixty-four per cent of the less retarded group and 61% of the severely and profoundly retarded group reached asymptotic acquisition described by these four patterns of rate indices.

Those who demonstrated acquisition and those who did not were not significantly different with respect to psychometric classification, age, age at admission, duration of institutionalization, or sex. Furthermore, neither long-term display of disrupting behavior (Barrett, 1971) nor psychometric "testability" was related to participants' acquisition. In fact, 28% of those who showed acquisition had been described as "untestable," and 52% of the "untestables" showed acquisition of one state or another. Of those, 36% reached full acquisition without special training.

In Figure 2, the distribution of severely and profoundly retarded participants among the six subgroups is compared with the distribution of the moderate to borderline group. The most striking feature of the two distributions is their overlap. Although there are notable differences in the first, fifth, and sixth subgroups, the distribution among acquisition subgroups attained by the severely and profoundly retarded group is not significantly different from that achieved by the moderate to borderline group. In other words, the amount of "learning"

TABLE 1

OPERATIONAL AND FUNCTIONAL DESCRIPTIONS OF ACQUISITION CRITERIA

PROGRAM OPERATIONS	STAGES OF ACQUISITION CRITERIA	DEMONSTRATED PROGRAM FUNCTIONS
<p>THE APPARATUS IS MADE AVAILABLE FOR REPEATED 60-MIN. SESSIONS UNTIL NO FURTHER BEHAVIOR CHANGE IS OBSERVED.</p>	<p><u>LOW RATE NONDIFFERENTIAL BEHAVIOR</u>: RATES OF ALL FOUR COMBINATIONS SO LOW THAT NO PATTERN IS SHOWN.</p>	<p>NO FUNCTIONS DEMONSTRATED.</p>
<p>LEFT LIGHT ILLUMINATED FOR ALTERNATING 60-SEC. PERIODS DURING WHICH EVERY 10TH OPERATION OF LEFT PLUNGER PRODUCES A CANDY, PENNY OR TOKEN.</p>	<p><u>NONDIFFERENTIAL BEHAVIOR</u>: INCREASED RATES OF ALL FOUR COMBINATIONS BUT NO CLEAR DIFFERENCES AMONG THEM.</p> <p><u>DIFFERENTIATION OF PLUNGERS</u>: WHILE LEFT LIGHT IS ON, S'S RATE OF OPERATING LEFT PLUNGER INCREASES AND RATE ON RIGHT PLUNGER DECREASES UNTIL THE DIFFERENCE BETWEEN THESE TWO IS 80% OR MORE OF THEIR TOTAL FOR EACH OF FOUR CONSECUTIVE SESSIONS.</p>	<p>WHILE PROGRAMMED CONSEQUENCE MAY BE FUNCTIONING AS A REINFORCER, THE DIFFERENTIAL REINFORCEMENT COMPONENTS OF THE PROGRAM HAVE NO FUNCTIONS FOR S.</p> <p>THE FUNCTION OF EACH PLUNGER HAS BEEN DIFFERENTIATED. S HAS LEARNED WHAT TO DO.</p>
<p>RIGHT LIGHT ILLUMINATED FOR ALTERNATING 60-SEC. PERIODS DURING WHICH NO REINFORCEMENT CAN BE OBTAINED.</p>	<p><u>DISCRIMINATION OF LIGHTS</u>: WHILE RIGHT LIGHT IS ON, S'S RATE OF OPERATING LEFT PLUNGER DECREASES RELATIVE TO S'S RATE ON THE SAME PLUNGER WHEN LEFT LIGHT IS ON UNTIL THE DIFFERENCE BETWEEN THE TWO RATES IS 30% OR MORE OF THEIR TOTAL FOR EACH OF FOUR CONSECUTIVE SESSIONS.</p> <p><u>CONTROL OF EVERGENERALIZATION</u>: WHILE RIGHT LIGHT IS ON, RATE OF OPERATING RIGHT PLUNGER DECREASES WHILE REINFORCED COMBINATION INCREASES UNTIL THE DIFFERENCE BETWEEN THESE RATES IS 80% OR MORE OF THEIR TOTAL FOR EACH OF FOUR CONSECUTIVE SESSIONS.</p>	<p>THE REINFORCEMENT-SIGNALING FUNCTION OF THE LEFT LIGHT HAS BEEN DISCRIMINATED FROM THE NONREINFORCEMENT-SIGNALING FUNCTION OF THE RIGHT LIGHT. S HAS LEARNED WHEN TO RESPOND.</p> <p>THE NONREINFORCEMENT-SIGNALING FUNCTION OF THE RIGHT LIGHT APPLIED TO BOTH PLUNGERS.</p>
	<p><u>FULL ACQUISITION</u>: WHILE LEFT LIGHT IS ON, S PULLS LEFT PLUNGER FOR REINFORCEMENT AND WHEN RIGHT LIGHT IS ON, S RARELY OPERATES THE APPARATUS. THE DIFFERENCE BETWEEN THE RATE OF OPERATING THE LEFT PLUNGER WHEN LEFT LIGHT IS ON AND RATES OF EACH OF THE OTHER THREE COMBINATIONS IS 80% OR MORE OF THE TOTAL OF EACH FOR FOUR CONSECUTIVE SESSIONS.</p>	<p>THE FUNCTIONS OF ALL PROGRAMMED CONTINGENCIES ARE DEMONSTRATED IN S'S OPERATION OF THE APPARATUS.</p>

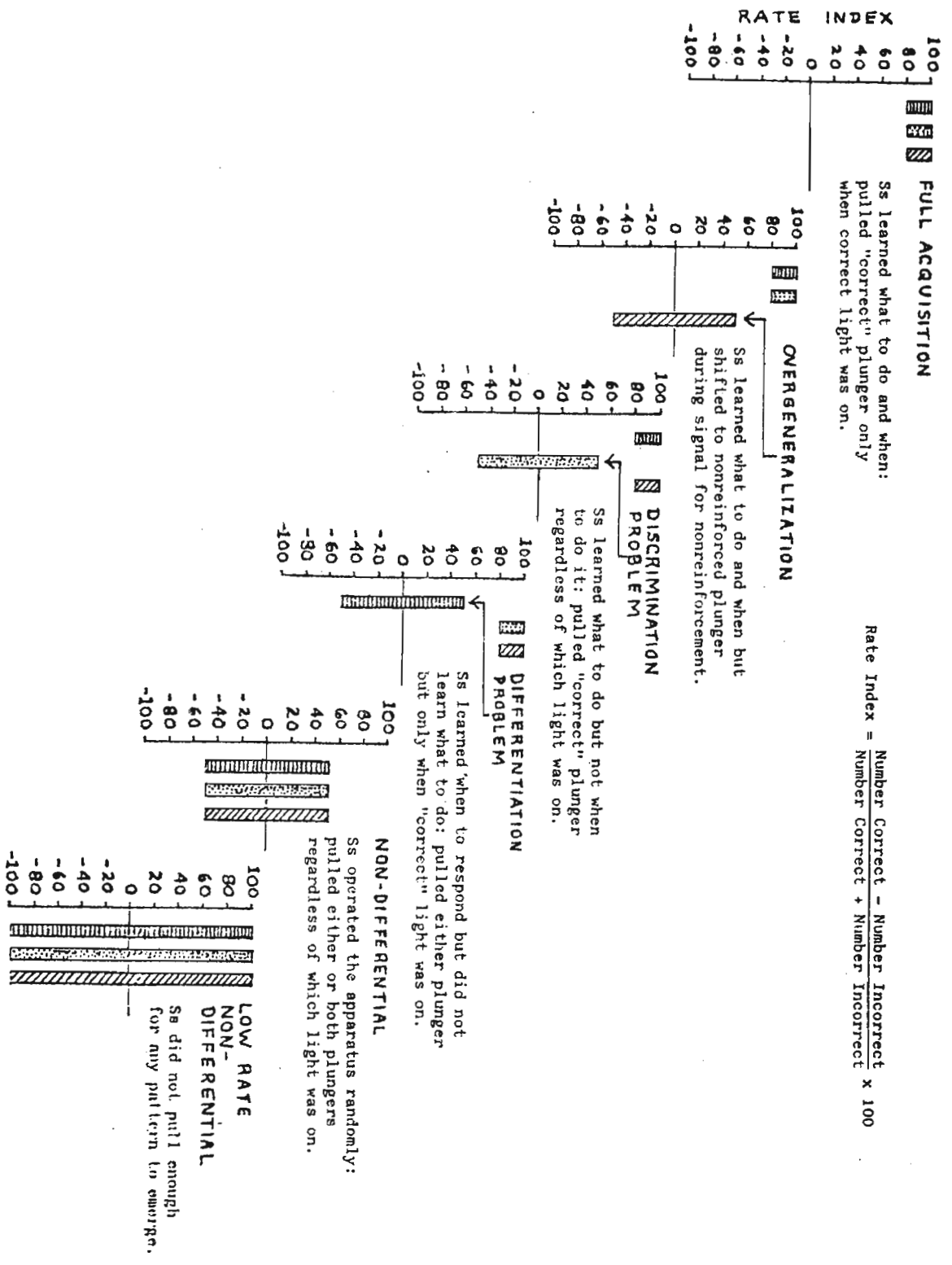


FIGURE 1

RANGES OF RATE INDEX PATTERNS THAT DEFINE EMERGENT ACQUISITION SUBGROUPS

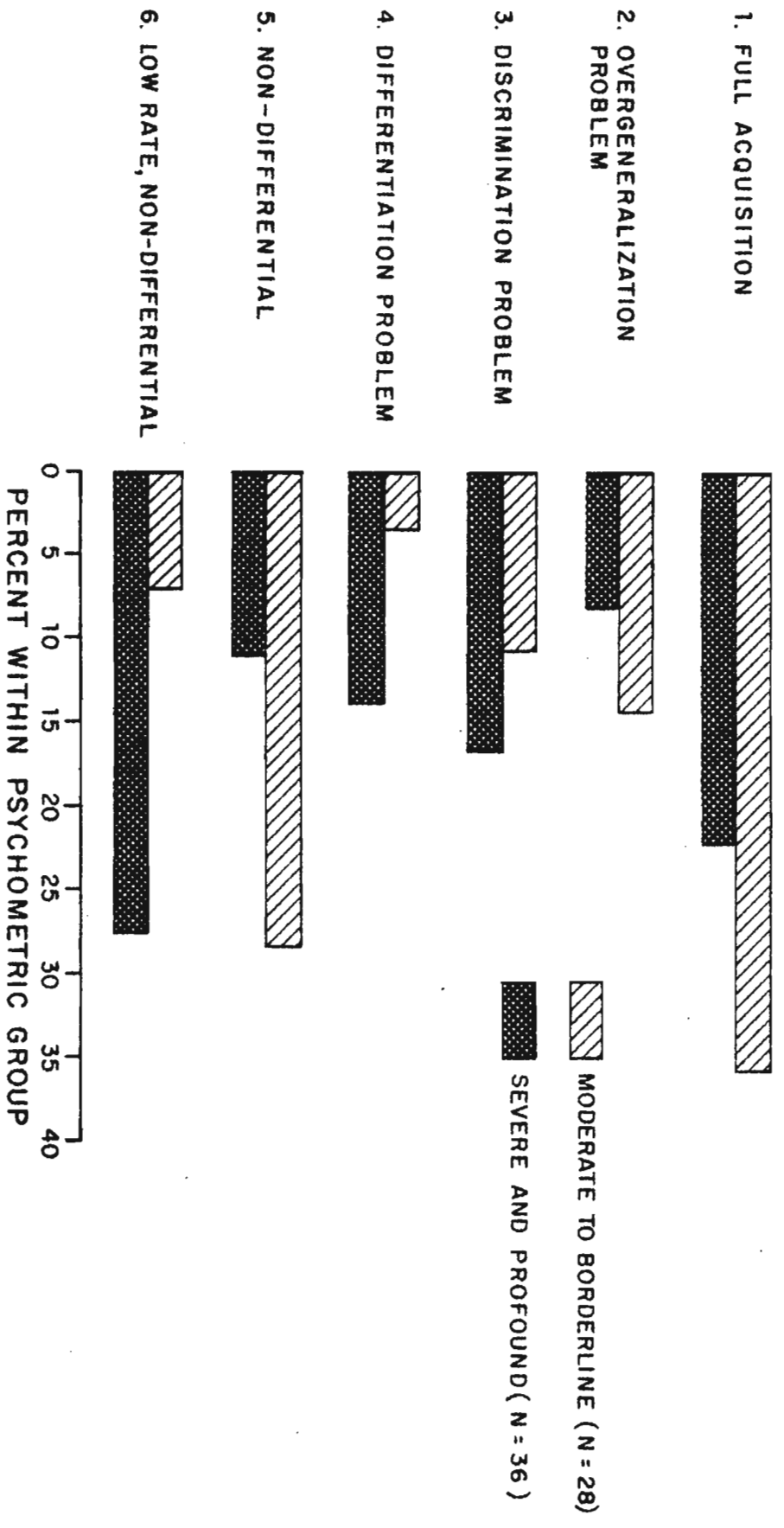


FIGURE 2

OVERLAPPING DISTRIBUTIONS OF PSYCHOMETRICALLY CATEGORIZED RESIDENTS
IN FINAL ACQUISITION SUBGROUPS

was not significantly related to psychometric level. Neither was it related to sex, age, age at admission, or duration of residency. Under the contingencies provided, "hopeless" residents learned as much as those considered "trainable" or "educable."

Nevertheless, the two psychometrically categorized groups did show differences, some of which have habilitative relevance.

As shown in Figure 3, for example, the groups differed significantly in the mean number of sessions to reach best performance ($p=.03$) and in the way they were distributed over acquisition time intervals ($p=.006$). Of the less retarded group, 65% reached peak performance within 10 sessions. The severely and profoundly retarded participants were more evenly distributed over acquisition time intervals, with 78% of the group taking more than 10 sessions to reach their best levels. Clearly, with severely and profoundly retarded people, we must be prepared to offer repeated learning opportunities over a much longer time than is usually necessary with less handicapped people.

Amount of time to reach best performance was not related to sex, to age, or to the degree of performance efficiency. But, irrespective of psychometric status, those who showed their best performance in the least amount of time were older at admission ($p<.01$) and had been at Fernald a shorter time ($p<.05$). For more accurate assessment of "trainability" a greater number of learning sessions was necessary for those who had been in the institution longer and from an earlier age. This finding is not surprising in view of the mounting evidence of deleterious effects of institutionalization, especially on children admitted when they are very young.

Another difference between psychometrically categorized groups is shown in Figure 4. If we look at the absolute rates of responding rather than rate indices (relative rates), we find that nearly 70% of the severely and profoundly retarded participants operated the apparatus at a median rate below 20 per minute during consecutive sessions at peak efficiency, whereas 64% of the less retarded participants responded at median rates of 20 per minute or above. At peak acquisition, the moderate to borderline residents operated the apparatus at much higher rates (mid-median: 23.5 times per minute) than the severely and profoundly retarded residents (mid-median: 11.0 times per minute) -- a highly significant difference ($p=.008$).

Rate of operating the apparatus was positively related to age ($.01<p<.02$) and to age at admission ($p<.001$). The rate was inversely related to duration of residency, but the relationship is not statistically significant. Thus, at peak acquisition, the older, less retarded residents who were older at admission emitted more behavior than those who were younger, more retarded, and had spent less time outside the institution.

We thought that high-rate participants might distribute themselves differently in the acquisition subgroups than the low-rate participants. However, if we discount the individuals in the "low-rate, nondifferential" group, for whom the procedure was obviously not appropriate, we find no significant difference. If a person operated the plungers at all, the rate was not related to how much he or she learned.

These findings have several implications for the design of assessment procedures appropriate for individuals within the severely and profoundly retarded range.

- Repeated samples of performance in a given situation are necessary to obtain a reasonable estimate of what a particular retarded person can do. A greater number of repetitions is necessary for assessment of severely and profoundly retarded people.
- Behavioral subgroups defined by commonalities in patterns of interaction with a standardized environment crosscut the psychometric categories that have homogenized the retarded population, especially those in the severely and profoundly retarded category. To reveal the individual patterns of abilities and deficits of persons who are beyond the sensitivity of conventional measurement techniques, behavior-assessing methods should permit a variety of "solutions" to the same problem.
- To assess learning potential, we must provide opportunities for learning (behavior change) to occur in response to clearly defined and therefore replicatable contingencies.
- Whatever the problem and however many the patterns of individual response that are possible, a standard measuring system should be used so that we can directly compare individuals throughout the clinical population of interest.

Acquisition Subgroups at Successive Stages of Emergence

We examined more closely the progressive changes in distributions among acquisition states of the 40 subjects who eventually reached States 1 to 4, as shown in Figure 2. In comparing the courses of acquisition shown by the two psychometrically defined groups, we asked two principal questions: 1) Do the distributions of the two groups among acquisition states differ at any successive stage, i.e., do they start off looking different and then begin to look more alike as more individuals reach asymptote? 2) What percentage of eventual "learners" appear as "nonlearners" at successive stages? Or, put another way, if we had set arbitrary time limits, what percentage of learners would have been misclassified?

The 40 subjects were 62.5% of our total group of 64. Eighteen (45%) of the 40 were classified as Levels III-V (borderline to moderate retardation), and 22 (55%) as Levels I and II (severe and profound retardation).

In this analysis, we considered repeated opportunities in the differentially reinforcing environment as repeated opportunities for individuals to be "sifted" into the various acquisition subgroups. For each individual, sifting was completed only when maximally accurate performance had been maintained for four consecutive one-hour sessions. The accuracy distributions of the two psychometrically defined groups were compared at the end of successive blocks of five sessions throughout the time necessary for all participants to show maximally stable rate index patterns. Once sifted into a subgroup, individual participants remained in the subgroup throughout analysis of subsequent session blocks. Thus, the percentages from distribution to distribution are cumulative, and the number of individuals remains the same throughout. Those individuals who were still in the process of reaching maximal acquisition during a given block of five sessions were assigned to an acquisition subgroup on the basis of their median rate indices for the session block.

Figure 5 shows the block-by-block distributions of the two psychometric groups as the individuals within each group sifted themselves according to their own

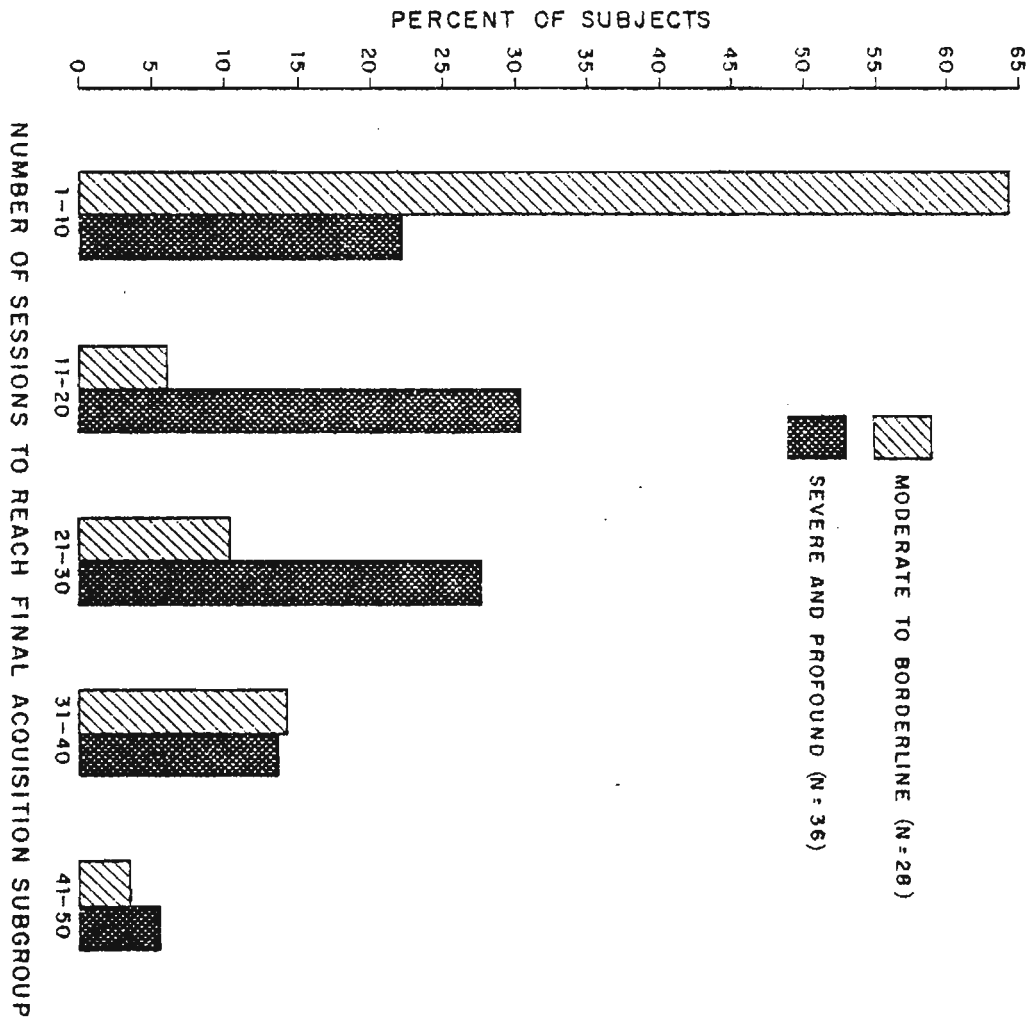


FIGURE 3

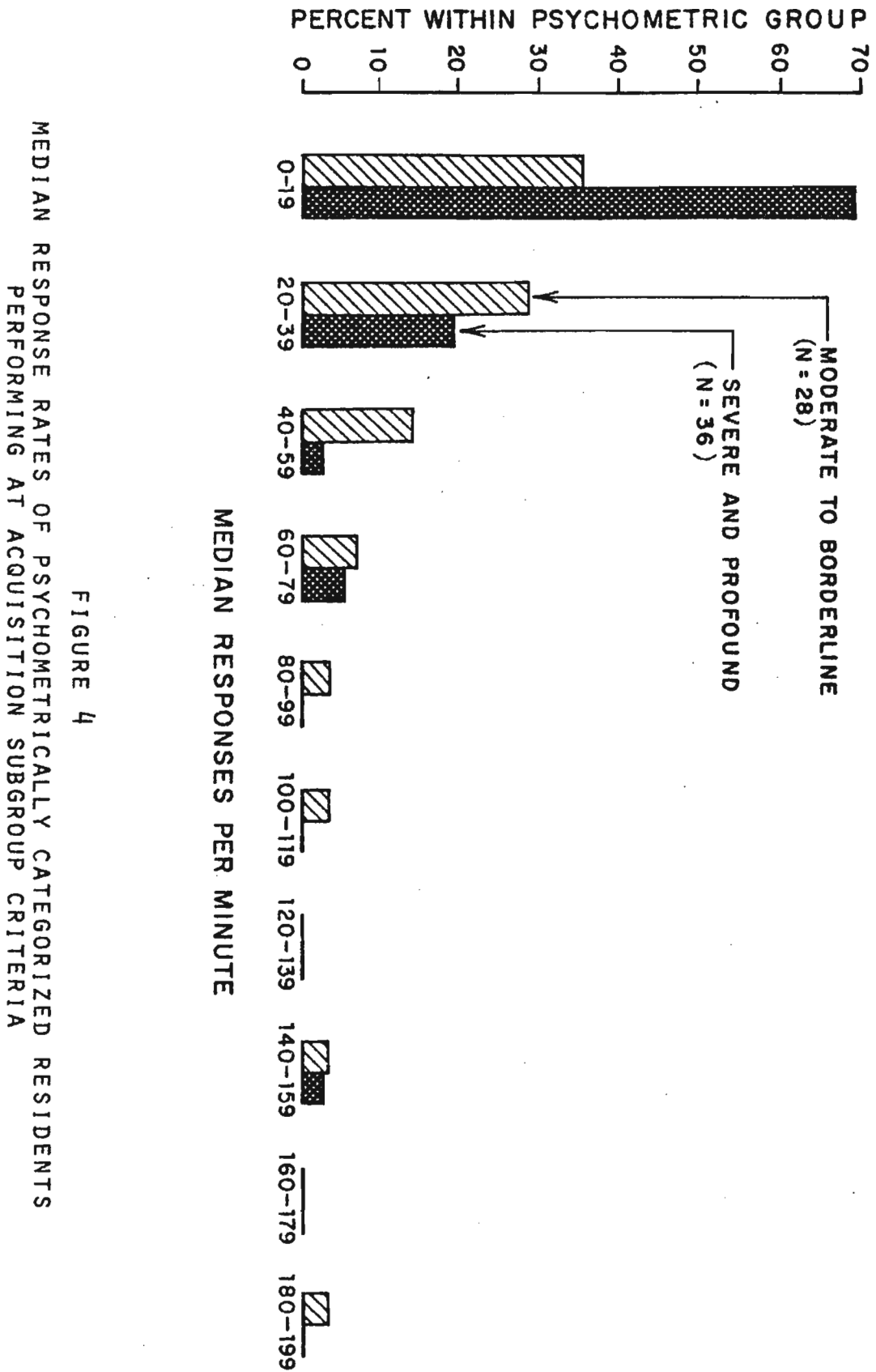
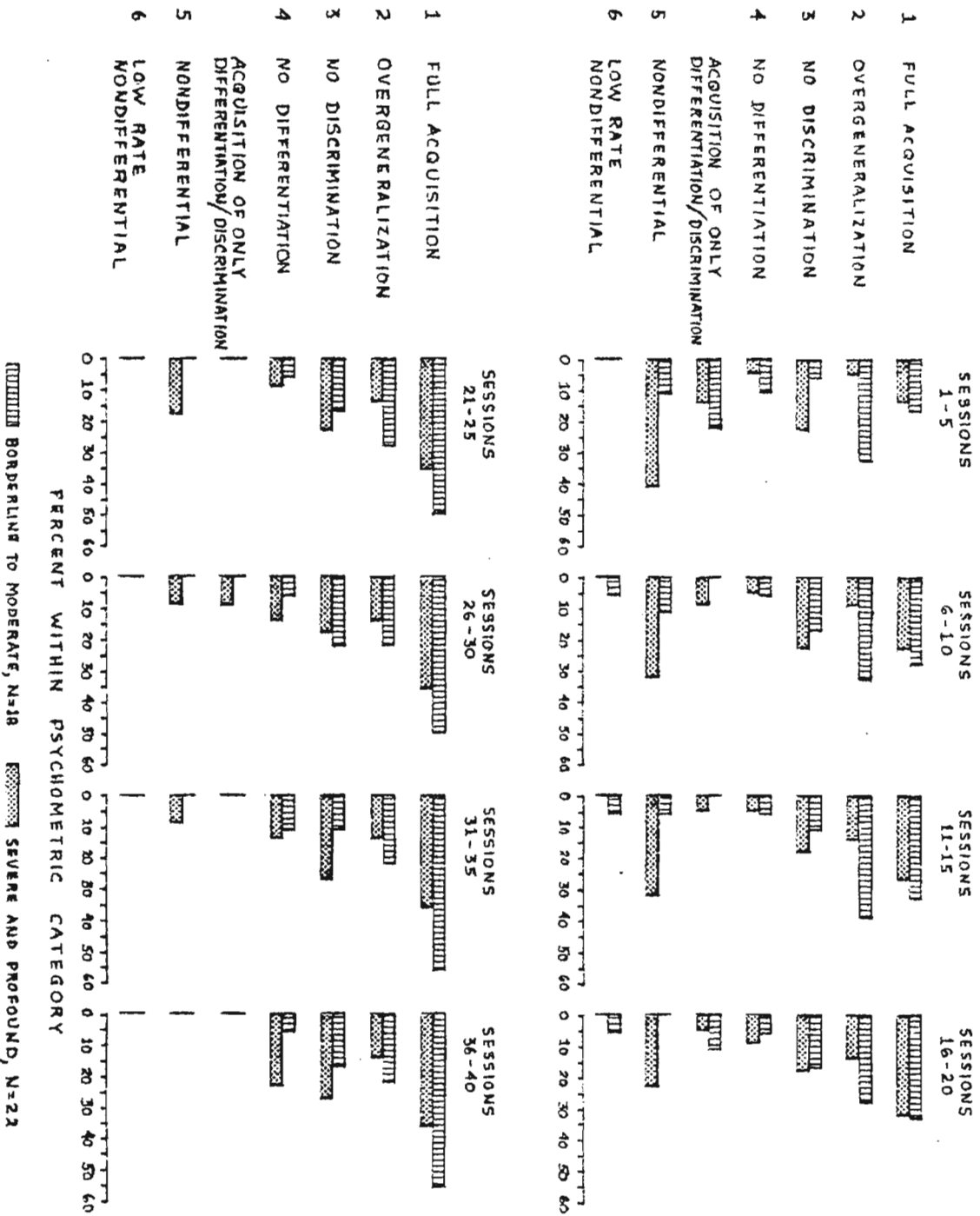


FIGURE 4

MEDIAN RESPONSE RATES OF PSYCHOMETRICALLY CATEGORIZED RESIDENTS
 PERFORMING AT ACQUISITION SUBGROUP CRITERIA

FIGURE 5
 STATUS OF ACQUISITION SUBGROUPS
 AT SUCCESSIVE STAGES OF EMERGENCE



"best acquisition states." The first segment of the figure shows how the 40 "learners" were distributed following the first five one-hour opportunities. The final segment shows how they were distributed among acquisition subgroups after every participant had been given repeated opportunities necessary for each to show his or her maximally accurate acquisition.

In no case are the distributions within session blocks significantly different for the two psychometrically defined groups. That is, with repeated opportunities to attain their best level of acquisition, the severely and profoundly retarded participants were distributed similarly to the borderline to moderately retarded participants. This is consistent with the comparison of final state distributions for all subjects shown in Figure 2. Degree of retardation, as defined by psychometric scores, appears to have little bearing on how much a person learns in (how efficiently he or she adapts to) our standard differentially reinforcing environment. This finding supports a growing body of evidence that predictions about retarded people's competency limits based on their current IQ or SQ are shaky at best and probably should never be made (see Gold, 1973).

The data in Figure 5 have even more important implications for design of reliable assessment methods. By comparing the successive distributions, we can trace the shifts of psychometrically categorized participants from transitional states to their best performance. Since we excluded from this analysis the 37.5% of our total group who showed no acquisition without special training, no one shows up in States 5 or 6 of the last distribution (sessions 36-40). The reason the additional state appears without a final-state number is that in this particular experimental design, acquisition of only the differentiation of the plungers or only the discrimination of the light positions is a transitional stage in the emergence of more enduring behavior patterns (Barrett, 1965).

Since it is common practice to set arbitrary limits on the number of behavior samples obtained for most assessment purposes (usually only one!), Table 2 summarizes the data from Figure 5 in terms of the percent of subjects showing any stage of acquisition after an arbitrary but specified number of repeated one-hour opportunities. The columns labeled "% False Negatives" indicate the percentage of subjects who might have been misclassified as "nonlearners" after an arbitrary number of sessions because their full performance efficiency had not yet emerged. If we had set a limit of five one-hour sessions, there would have been 28% false negatives in the total group, 11% among the less retarded and 41% among the more severely retarded. After 25 hours there would have been no false negatives among the less retarded but still 18% among the more severely retarded. After 40 sessions, all subjects showed some form of acquisition. Fifty-five percent of the borderline to moderately retarded participants and 36.4% of the severely and profoundly retarded participants achieved State 1 (full acquisition); 44.4% of the borderline to moderately retarded people and 63.6% of the severely and profoundly retarded people were identified as showing one of the specific deficits (States 2-4) requiring special training.

Some may regard these data as resulting from a repeated "trial-and-error" procedure with intermittent reinforcement and without any added remedial or prosthetic variables, and they may register astonishment that anyone showed acquisition. The more psychometrically minded may think of repeated testing to sort people into groups that are homogeneous with respect to degree of acquisition or type of deficit. The experimental analysts of free operant behavior may see in these data the development of baseline performances prior to experimentation (the clinical analog would be a kind of "waiting list cure"). We hope that professionals who make decisions about retarded people's adaptability to or

acquisition in novel environments will see their own analogs in these data and, at the very least, pause to ask whether, indeed, they have permitted retarded persons the repeated opportunities they need to show what they can do and what specific problems they need help with.

Our own observations of our classroom pupils and our vocational trainees — as well as the results of repeated psychometric testing, even without added reinforcement (Ayllon and Kelly, 1972; Edlund, 1972) — are corroborated not only by our laboratory data but also by the reports of other investigators who find that, regardless of whether they are living at home or in an institution (Bricker, 1972), retarded persons require a minimum adaptation period of at least six weeks before anything resembling reliable samples of current skills can be obtained.

Yet, in the absence of any normative information on the time required for maximal adaptation to varieties of situations (to a workshop situation, to a new ward environment, etc.), pronouncements are made and habilitative decisions are reached about individuals on the basis of an inadequate number of observations over an even less adequate period of time. At the very least, this practice creates artifacts in evaluating the effectiveness of training procedures. Most importantly, it increases the likelihood of failure of the very mission it seeks to accomplish: to maximize the opportunities and options for retarded persons to perform more normally.

Rate Distributions During Acquisition

This analysis focuses on reinforced response rates, rather than total amount of responding, that accompanied successive stages of acquisition. Were there differences between the two psychometric groups? If so, when did the differences first appear?

Figure 6 compares the two psychometrically defined groups in terms of how rapidly they were doing what they got reinforcement for (that is, operating the left plunger only when the left light was on). Individual medians are based on the same blocks of five sessions that we looked at for successive accuracy comparisons. Median tests comparing the reinforced response rates of the severely and profoundly retarded with those of the less retarded yielded p values of .03, .06, .05, .03, .0067, .0067, .0067, .0067 for the successive blocks of sessions. As more individuals reached maximum accuracy (that is, as more stable accuracy patterns emerged), the rate differences between the two groups yielded successively lower p values. The overall relation of rate to psychometric category yielded p less than .001.

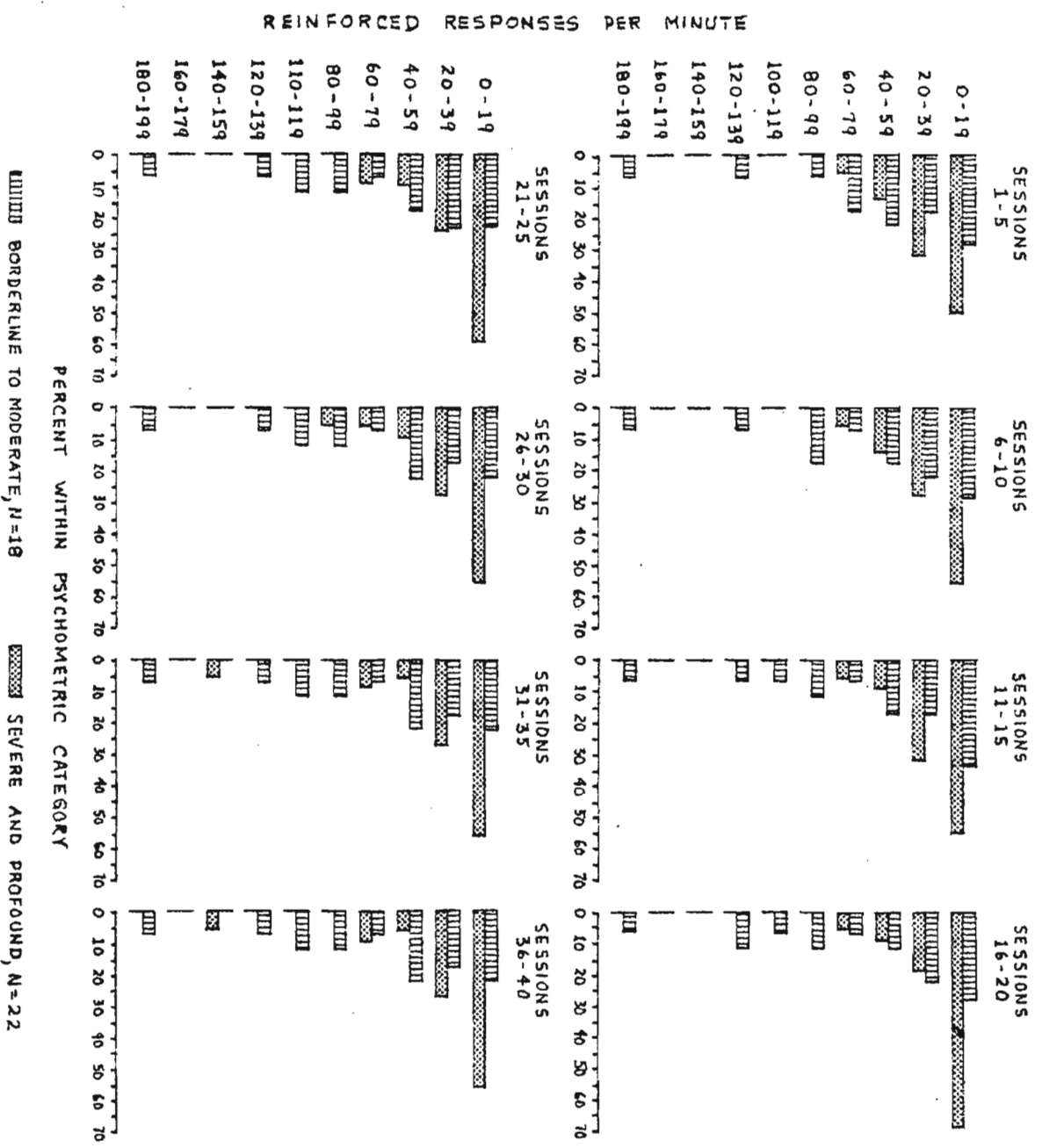
This comparison of psychometrically defined groups suggests that, in a differentially reinforcing environment, severity of retardation may not determine the upper limits of accuracy. But without procedures designed to increase behavior frequency, no matter how accurately a person performs, the more severely retarded will still show a substantially greater rate deficit than the less retarded.

Rates of reinforced responding were not significantly related to age, sex, or degree of acquisition. They were significantly related to age at admission and duration of residency. The older a person was when admitted, the higher his or her reinforced rate (p less than .01). And the longer the person was an institutional resident, the lower his or her rate of reinforced behavior ($p=.048$). These findings are not surprising in view of another study (Barrett, 1971) which

Table 2
 Changes in Acquisition Status Shown by Repeated Measures

After	Total Group		Borderline to Moderately Retarded		Severely and Profoundly Retarded	
	% Showing Any Acquisition	% False Negatives	% Showing Any Acquisition	% False Negatives	% Showing Any Acquisition	% False Negatives
5 sessions	72	28	89	11	59	41
10 sessions	75	25	83	17	68	32
15 sessions	77	23	89	11	68	32
20 sessions	85	15	94	6	77	23
25 sessions	90	10	100	0	82	18
30 sessions	95	5	100	0	91	9
35 sessions	95	5	100	0	91	9
40 sessions	100	0	100	0	100	0

FIGURE 6
RATE DISTRIBUTIONS ACCOMPANYING EMERGENCE OF ACQUISITION SUBGROUPS



suggested that institutions may, inadvertently, differentially reinforce low rates. We found that more heavily staffed and more protective living environments were associated with lower rates of performance.

Clearly assessment should ask not only "How well can the individual do something?" but also "How rapidly?" If only accuracy of performance is taught, severely and profoundly retarded persons will continue to exhibit behavioral retardation, for they will not have learned to perform accurately at a more normal rate. Behavioral normalization must consider multiple dimensions of behavior to produce a credible product.

To avoid misclassifying individuals with more severe behavioral retardation and thus inadvertently limit their opportunities for habilitation, repeated measures should be obtained. This practice appears to be especially important for individuals who were admitted to an institution at young ages and for those who have been in an institution for a long time.

References

- Ayllon, T., and Kelly, K. Effects of reinforcement on standardized test performance. Journal of Applied Behavior Analysis, 1972, 5, 477-484.
- Barrett, B. H. Acquisition of operant differentiation and discrimination by institutionalized retarded children. American Journal of Orthopsychiatry, 1965, 35, 862-885.
- Barrett, B. H. Behavior individuality in four cultural-familially retarded brothers. Behaviour Research and Therapy, 1969, 7, 79-91.
- Barrett, B. H. Behavioral differences among an institution's back ward residents. Mental Retardation, 1971, 9(1), 4-9.
- Ericker, W. Personal communication, July, 1972.
- Edlund, C. V. The effect on the test behavior of children, as reflected in the I.Q. scores, when reinforced after each correct response. Journal of Applied Behavior Analysis, 1972, 5, 315-319.
- Gold, M. W. Research on the vocational habilitation of the retarded: The present, the future. In N. R. Ellis (Ed.), International review of research in mental retardation (Vol. 6). New York: Academic Press, 1973. Pp. 97-147.
- Heber, R. A manual on terminology and classification in mental retardation. (American Association on Mental Deficiency.) American Journal of Mental Deficiency, monograph supplement, 1959, 64. (No. 2)