DOCUMENT RESUME

ED 059 274 TM 001 091

AUTHOR Cline, Marvin G.
TITLE Stanford-Binet, PPVT, and Low-Income Preschoolers: New Pitfalls for Old Tests.
INSTITUTION Howard Univ., Washington, D.C.
PUB DATE Apr 66
NOTE 16p.; Paper presented at the Annual Meeting of the Eastern Psychological Association, New York, New York, April 1966

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Academic Performance; Age Differences; Anxiety; *Cognitive Development; Disadvantaged Youth; Educational Problems; *Language Development; Language Skills; *Low Income; Measurement Instruments; Negro Students; Norm Referenced Tests; Oral Expression; Poverty Research; *Preschool Children; Primary Grades; Program Evaluation; Receptive Language; Self Concept; Sex Differences; *Testing Problems
IDENTIFIERS *Peabody Picture Vocabulary Test; Stanford Binet

ABSTRACT

To reemphasize the dangers inherent in the misuse of testing instruments, a study of language and cognitive development in poverty preschoolers investigated 1) whether the interpretation of Peabody scores could be applied to this population, and 2) the contribution of the linguistic form of the Peabody to performance. (MS)

Stanford-Binet, PPVT, and Low-Income Preschoolers: New Pitfalls For Old Tests*

by Marvin G. Cline**

One of the consequences of almost every war is the development of a new technology for the creation of weapons of destruction. Occasionally the new technology gives us the opportunity to beat our swords into plowshares with somewhat greater efficiency than before the war. This is not meant, of course as a justification of war. It is, rather, a statement of expectation, of what might occur as the result of the currenr battle against poverty. Unfortunately, except for a few minor variations in tactics, the same research strategies with the same techniques seem to be found in most of the battles of this war. In this case, we refer to the standard technique of putting easy-to-use "quickie" tests in the hands of ill prepared examiners (teachers, teacher aides, housewife volunteers, nurses, etc.), to administer to non-normative population. The battle is the identification, evaluation and elminiation of educational problems of the preschool and primary grade low-income child. During the past 20 years, a continuous warning has been given to administer instruments only to populations on whom the instruments were standardized, to be aware of the contribution of the testing situation to test performance, and to consider the self-defeating strategies which low-income children use in testing situations. Nevertheless, the decision to use an instrument such as the Peabody Picture Vocabulary Test to diagnose low-income preschool children, and to evaluate the

---

* Read at the Annual Meetings of the Eastern Psychological Association, New York City, April 14-16, 1966.

** Judith Marshall, Eunice Stansbury, Howard University, Washington, D.C.

effects of nursery school participation in such large-scale programs as Head Start, clearly ignore these warnings.

The dangers inherent in such procedures are large and apparent: First, the understanding of the impact of poverty on the development of the child is based on the identification of the performance deficit in the first place. Hypothetical explanations of the intellectual deficit of the children of the poor range from test anxiety to restricted sensory stimulation. The former category assumes that the observed deficit is an artifact. The latter explanation assumes that the deficit is an accurate picture of the level of development, and is more or less reversible, depending upon whether the theorist has worked with infrahumans or adolescents. Clearly the data required for the evaluation of these hypotheses are the nature and extent of the deficits elicited on measurement.

The second danger is the evaluations of experimental programs will be so filled with extraneous variables as to render the whole effort worthless. Worse yet, policy decisions for the continuance or discontinuance of programs may be made on the basis of these evaluations, and the plaintive wails of the social scientist to beware that spurious data will be lost in the storm of spending for more of the same or more of something new.

In order to re-emphasize the dangers inherent in these procedures, a small study of language and cognitive development in poverty preschoolers was devoted to one aspect of the standard testing of our subjects. We decided to administer the Peabody along with the Stanford-Binet in order to determine if the interpretation of Peabody scores, which is partly justified by the large amount of variance shared by these two measures, can be applied to

3

this population.

Next, we decided to investigate the contribution of the linguistic form of the Peabody to performance. It is reasonable to assume that the low-income Negro child will have some strong reservations about being tested, if the current conceptions about the negative self-image of the poverty child hold true. More important, however, are the responses these children might have to a measure of a language which is completely passive. Recall that the tasks of the Peabody is to listen to the word spoken by the examiner and to point to the picture on the test booklet which best represents that word. This is measuring <u>receptive</u> language, i.e., understanding the speech of others. It does not measure expressive language, which involves the free utilization of verbal symobls for communication purposes. One may understand the speech of another without the ability to speak that language (e.g., animals and preverbal children), although we ordinarily assume that when we attempt to control the behavior of another via verbal communication, the other has an adequate understanding (receptive language) of our expressive language. Clearly, these two verbal systems have different functions in the social, and therefore, testing behavior of subjects. Receptive language is controlled by the other, whereas expressive language attempts to control the other. We have reasoned that in a stressful situation (and we are assuming that testing situations are stressful for those children who own the legacy of second-class citizenship), to be controlled is more threatening and therefore more debilitating, than to be the controller. In this case, to be presented with a single response, all or none situation (i.e., the receptive form of the Peabody), weakens the defense against the stress of evaluation. We are assuming that the need to defend

against the threat implied in being evaluated is very strong in lower income Negro children. Here, the child cannot explain or justify his choices, he has no way of recording his hunches or partial information and he has no alternative but to point to a single picture. Just as with the high anxiety college sophomore who is ego-involved in a complex task, the low-income child can be expected to show maximum performance deficit via withdrawal and/or random behavior, under those conditions.

One approach to this problem is to construct an expressive form of the Peabody to compare with the scores from the receptive form. Predictions about discrepancies between these scores would depend upon three other factors as well as the assumed anxiety-proneness of these subjects. Since we are dealing with a verbal term, there should be some contribution of verbal skill to the tendency to make up in a less stressful verbal situation what is lost in a stressful one. Thus, we would predict that although the impact of stress is independent of initial verbal skill, the tendency to do better on an expressive (less stressful) form than on the receptive (stressful) form, is related to verbal skill. The second factor is the sex of the child, since a number of workers have noted significantly different patterns of verbal skills in low-income Negro boys and girls. The third factor is age of the child. There should be significant sex by age interactions in these data.

Since we are largely investigating the Peabody, it is not possible to use it as the measure of verbal skill. Consequently, we will assume that the Stanford-Binet is an unbiased measure of verbal abilities.

Subjects: Forty-six Negro children (24 boys and 22 girls), 46-68 months of age (median: 56 months) were randomly selected from a population of low-income children attending preschool centers run by the Community Action Program of Washington, D.C. The centers are all

located within the boundaries of the main target area of the anti-poverty program, and all children live in the neighborhood serviced by the center he or she attends. Admittance into the program was based upon financial need. The median family income of our sample, $3500 per year, is the same as the median for the total preschool population and the neighborhood. The median level of education of the parent is less than six years, again the same as the total preschool population and the total community.

Procedure: The Stanford-Binet (Form LM, 1960 revision) was administered to all children between the 3rd and 5th month of attendance. Starting the 7th month of attendance, a battery of linguistic measures was administered of which we shall report data from the following two:

a. Peabody Picture Vocabulary Test, Form A, administered, according to the standard procedure described in the test manual, by Negro female undergraduates who were trained specifically for this task. There were no experimenter effects apparent and the data from the different test administrators were combined into a single group. We call this the receptive form of the Peabody, since the child simply points to a picture on each place of four pictures which represents the word announced by the examiner.

b. Peabody Picture Vocabulary Test, Form B, administered in modified procedure by the same examiners. Plates 25-60 (representing the approximate range for this population) were used. They were presented to the child one at a time, with the examiner pointing to the picture on each plate that represented the word on the Form B list as described in the test manual. With each plate, the examiner pointed to the pre-selected picture and asked a standard question: "Tell me about this picture" or "What is in this picture?" The child

was allowed to speak as long as he wished about each picture, and his responses were recorded verbatim. Independent judges scored each response according to criteria, established and tested in advanced, which allowed for two points for use of the exact word or its synonym, one point for a functionally complete and correct description of the contents of the picture, and zero for an incorrect response. Total score is the sum of scores on each plate. Two independent judges scored each item with better than 95 percent agreement, and all disagreements were adjudicated in conference. Item analyses were carried out which led to the elimination of only two items. However, the data reported here are based on the original, unrefined form of this measure. The refined form is now being administered in a cross validation study. We call this the expressive form of the Peabody.

We are aware that the cognitive content of the two forms of the test are quite different, although the vocabulary lists are comparable. The child is not required to discriminate among the picture on the expressive form as he is on the receptive form, but since the vocabularies are comparable, a comparison of the relative skill each child shows on both test is a meaningful measure of similar verbal skills under different conditions. We have not balanced the order of pairing the two vocabulary lists with the two forms of the test, although this is being done in the replication study.

Results: The population was divided into boys and girls and those above and below the median age of 56 months. (Table 1)

The means Stanford-Binet IQ for the total population is 99.4, sigma 14, range 72-130. There are no differences between boys and girls, and a slight but not significant difference between those above and those below the median age: the IQ of the younger is 103.4 and

for the older group is 94.8. We consider this population comparable to the normative population of the same age. They are also comparable in IQ to the population used by Anastasi in her studies of preschool children using the Goodenough, and the population used by Deutsch using the Lorge-Thorndike. Apparently, when random samples of the lower income preschool populations are given carefully constructed intelligence tests, they do not show any meaningful deficit.

The receptive form of the Peabody shows rather a different picture. The mean Peabody IQ is 81.4, sigma 17.5, range 55-139. Further, the pattern of IQ scores across age and sex is interesting: younger girls and older boys (89 and 86 respectively) are significantly better than younger boys and older girls (77 and 74 respectively). (Table 1)

The relation between the Peabody receptive and the Stanford-Binet is our prime interest, however, and here we find large discrepancies between the two. (Table 2)

In all but four cases, subjects did better on the Stanford-Binet than on the Peabody. The mean discrepancy between the two is 17.5 IQ points in favor of the Stanford-Binet. If we divide the population at the median Stanford-Binet IQ of 100, and consider these discrepancies, there is a discernible pattern. The higher IQ subjects show a slightly greater deficit on the Peabody relative to the Stanford-Binet than the lower IQ subjects (21 IQ points in favor of S-B for the higher and 15 IQ points in favor of S-B for the lower). However, if we consider the extreme ends of the Stanford-Binet distribution, a significant (.01, Mann-Whitney one-tailed U test) difference emerges: those whose Stanford-Binet IQs are over 110 show a mean IQ reduction of the Peabody of 25.6 points, whereas those whose Stanford-Binet IQs are below 86 show a mean reduction 6.6 points. Three of the four

subjects who showed a higher Peabody IQ than Stanford-Binet IQ are in this lowest quartile of Stanford-Binet scores. The fact that the higher IQ subjects showed greater Peabody deficit than the lower, a phenomenon that looks something like a regression to the mean, did not produce a reduced variability in the Peabody scores. There was, however, a very small but significant correlation of .34 between the two sets of scores indicating some real problems in deciding what it is that the Peabody is measuring in this population.

Turning now to the expressive form of the Peabody, the older children have significantly higher raw scores than the younger children. (Table 1) This indicates that the instrument has some construct validity. However, the older girls, who showed the lowest receptive scores and the lowest Stanford-Binet IQs, do not show significantly better expressive scores than the younger girls. Older boys do show significantly better expressive scores than younger boys.

In order to compare the expressive with the receptive scores, each was transformed into standard scores. Comparison between scores will be in standard units hereafter. (Table 3)

The first question we asked of those data is whether or not our population did better on the expressive than the receptive. For the total population, the mean differences in Standard Scores should be zero, and this is what we found. However, we predicted that the expressive should show higher scores than the receptive, primarily in those who are verbally facile, and it is therefore necessary to group the population according to Stanford-Binet IQ scores. Breaking the population at the median reveals this tendency but not quite to significance; the higher IQ subjects show a mean of .30 standard scores discrepancy (expressive over receptive) whereas the lower IQ subjects show no difference between their expressive and receptive

scores. When we further break the population at the top and bottom quartiles, the tendency becomes significant. Those whose IQs are above 110 show a mean increment of the expressive over the receptive of .84 standard score points, whereas those below an IQ of 86 show an increment of the receptive over the expressive of .18 standard score points. A Mann-Whitney one-tailed test reveals this difference to be significant at the .025 level.

Further trends in these data: boys tend to do better on the expressive than the receptive, and the girls tend to do better on the receptive than the expressive; Ss below the median age tend to do better on the expressive than the receptive and those above the median age tend to do better on the receptive than the expressive. Neither of these trends is significant, but when the interaction is considered, a significant trend does appear. Young boys do better on the expressive than the receptive and older girls do better on the receptive than the expressive. This difference is significant (Mann-Whitney, one-tailed U test) at the .06 level. (Table 3)

We mentioned earlier that discrepancies between the Peabody receptive and the Stanford-Binet reflect the restrictive nature of the receptive task for this population. We also indicated that those with the most verbal skill should be able to recoup their losses via the expressive form of the test. Since all subjects except those in the lowest quartile of the Stanford-Binet IQ distribution showed very large losses in the Peabody receptive, it follows that those with large Peabody-Stanford-Binet discrepancies should also show large expressive-receptive discrepancies. This is what occurred. Those above the median Peabody-Stanford-Binet discrepancy of 19 points showed higher scores on the expressive than the receptive, and those below the median showed higher scores on the receptive

than the expressive. A Mann-Whitney one-tailed U test revealed this difference to be significant at the .025 level. (Table 4)

Discussion: The testing manual of the Peabody reports a correlation of .75 between it and the Stanford-Binet. It is this fact that undoubtedly led the test authors to have confidence in the concept of a Peabody IQ, although the large amount of variance the tests have in common includes the common chronological ages found in both IQs. Nevertheless, the Peabody is understood to be an estimate of intelligence, and a predictor of academic achievement because of this high correlation. Clearly it cannot have this function with the population of this study since, despite the common chronological age elements in both measures, the correlation is only .34. It is not clear what the Peabody is measuring.

Our tentative approach to this is to assume that the Peabody inhibits expression of verbal skills in the present population, and that Peabody scores should be lower than Stanford-Binet scores. This was strikingly apparent. However, our prediction that the deficit would occur at all levels of intelligence (Stanford-Binet) was not supported. The lowest quartile Stanford-Binet IQ subjects showed only a mean of 6.6 IQ points less on the Peabody than the Stanford-Binet, whereas all other subjects averaged 16-25 points less on the Peabody than the Stanford-Binet. However, the youngest boys and the oldest girls showed both the highest and the lowest Stanford-Binet IQs respectively, but they also showed the greatest negative discrepancies between the Peabody and the Stanford-Binet, but the girls do significantly better on the Peabody than the boys. Whatever it is that is depressing the Peabody scores relative to the Stanford-Binet IQs is distributed with an age by sex interaction effect which must be understood before any interpretation of the Peabody can be made.

Our attempt to explore this further by means of an expressive form of the test must be considered very tentative because the instrument has not been fully refined. However, even this crude form reveals the trends we predict: the higher IQ child, presumably the more verbally facile (in this population they tend to be the boys rather than the girls, and the younger rather than the older children), tend to do better on the expressive than on the receptive form. This appears to be similar to the improved performance shown by high anxious college students in situations of equal stress but less complex task demands. The behavior of Vera John's subjects (somewhat older but from the same socio-economic levels as the present population) when repeating a timed task with the time restrictions removed is another example of a similar dynamic.

We draw two conclusions from the present study:

1. The popular description of low-income preschoolers as mute and dull cannot be supported by these data. It is hard to see them as lacking in the necessary stimulation for normal verbal and intellectual development. They probably have more than their share of test anxieties and negative self-images and in the school situation where they will be treated in a manner that verifies their expectation of stress and failure, they will have more than their share of low scores on most tests. But there is little evidence that they are underdeveloped as yet. We shall leave that to the schools to accomplish.

2. It is imperative that large-scale replications of this kind of investigation take place to test the hypothesis that test anxiety contribute inordinately large amounts of variance to performance. Until this is done, current tables of norms are inappropriate for these populations.

This last is not a new notion. It is not even new that the structure against indiscrimination use of standard test on non-

12

normative populations is floured again and again. The only thing new here is that it is being done on such a grand scale; and in an area where the social science data and techniques are contributing to public policy.

Table 1

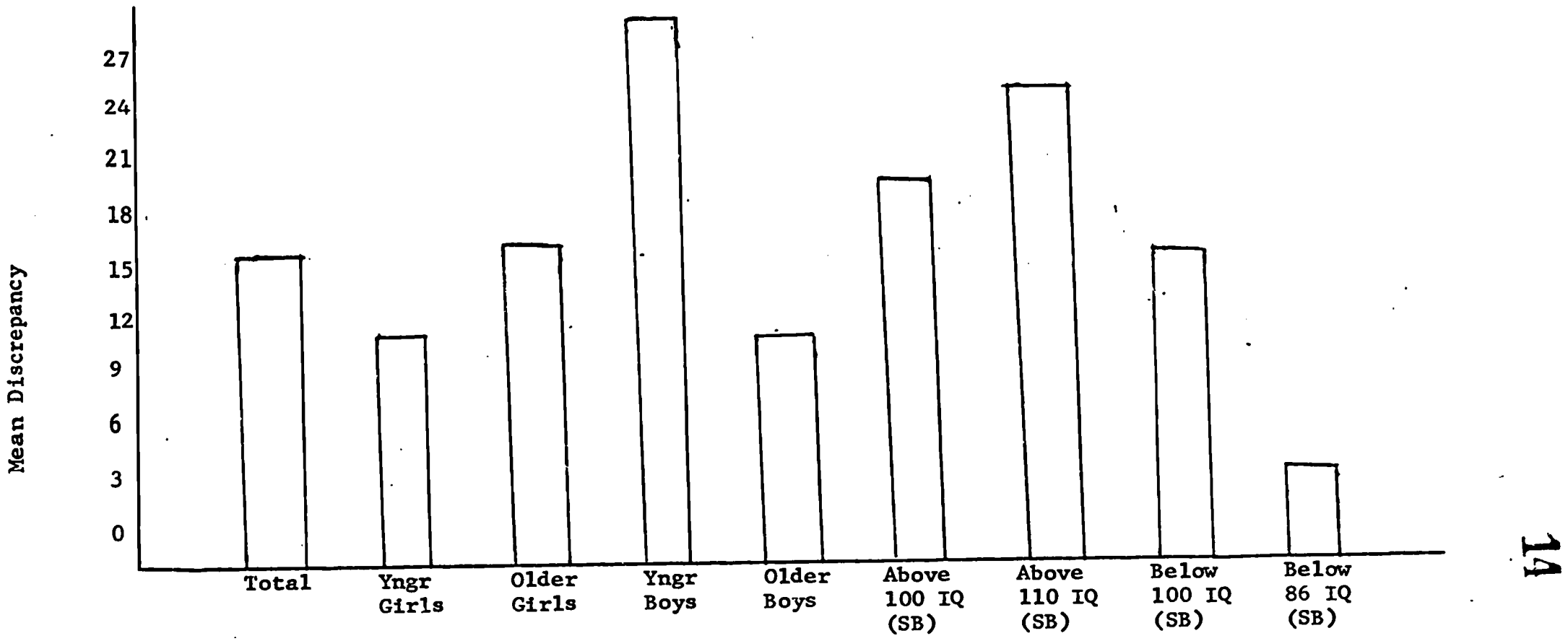
PPVT(R), STANFORD-BINET, PPVT(E) - Means and Standard Deviations

	PPVT			S-B		PPVT*	
	N	R	σ	IQ	σ	E	σ
Younger Girls	11	89.1	11.2	103.6	16.3	30.9	6.8
Older Girls	11	74.0	18.5	92.3	16.3	35.1	9.7
Younger Boys	13	76.7	14.0	104.4	9.5	31.0	6.5
Older Boys	11	86.0	20.5	97.5	12.9	44.3	5.5
Total Younger	24	82.6	14.3	103.4	13.1	31.0	6.2
Total Older	22	80.0	20.5	94.8	12.6	40.0	9.3
TOTAL	46	81.4	17.5	99.4	14.0	36.3	11.9

* Raw Scores

Table 2

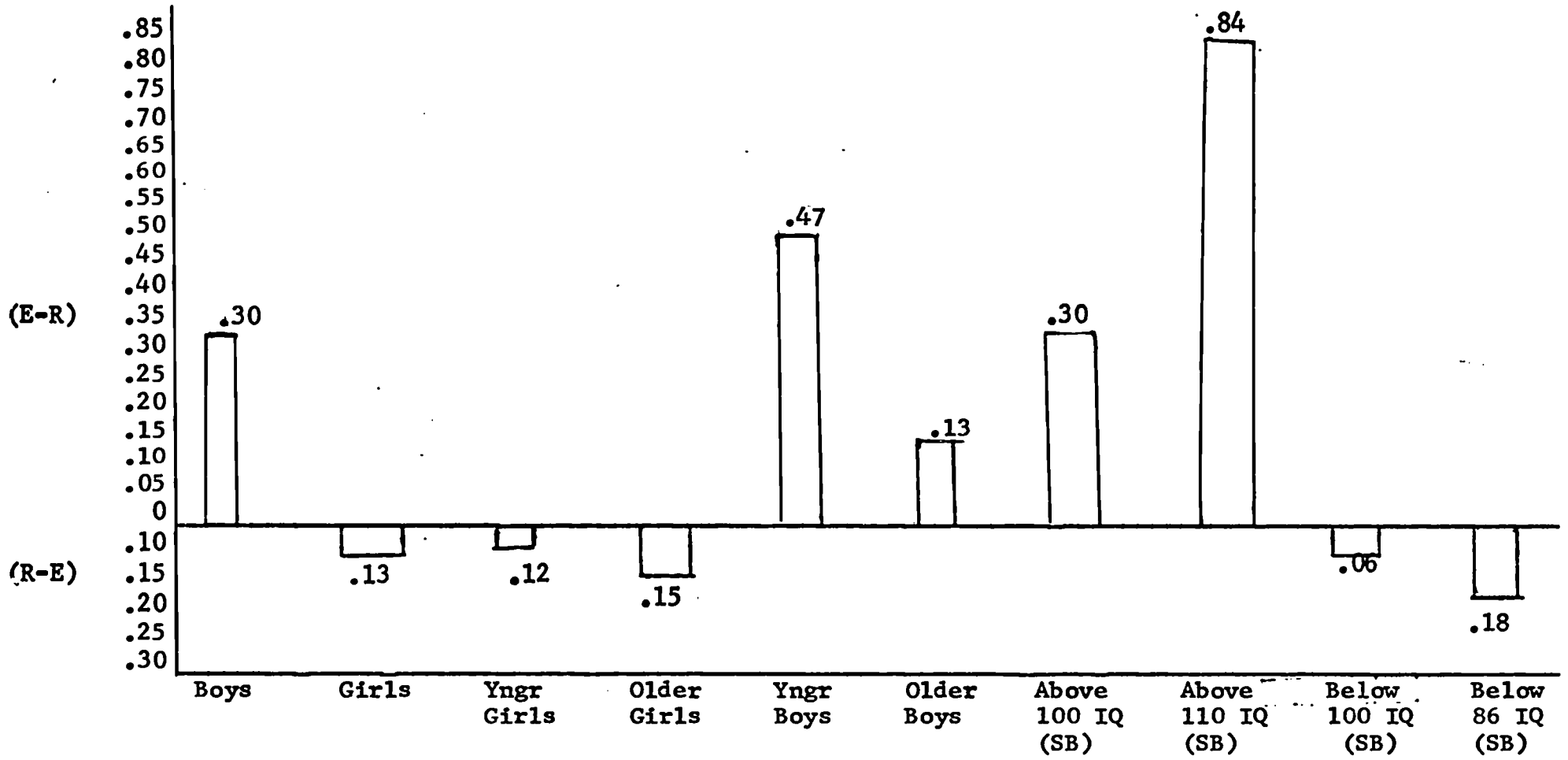
PPVT(R) - SB IQ - Mean Discrepancies (All Discrepancies in Favor of SB)



14

Table 3

PPVT R-E DISCREPANCIES*



* In Standard Scores

15

Table 4

PPVT R-E DISCREPANCIES X PPVT(R) - SB DISCREPANCIES

R - SB MEDIAN DIFFERENCE = 19 IQ POINTS

	ABOVE MEDIAN	BELOW MEDIAN
Number of Subjects w/Expressive higher than Receptive	15	7
Number of Subjects w/ Receptive higher than Expressive	8	15
Mean Discrepancy	.48 (E,R)	.33 (R,E)

16