DOCUMENT RESUME

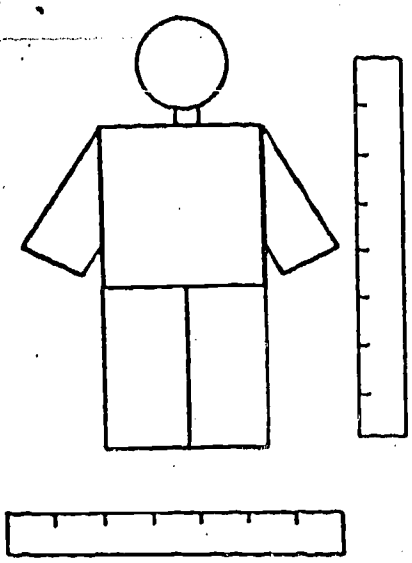
ED 135 868 TM 006 100

AUTHOR De Gracie, James S.; Vicino, Frank L.
TITLE A New Yardstick for Comparing Achievement in Basic Skills.
PUB DATE [Apr 77]
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Achievement Rating; Achievement Tests; Basic Skills; Criterion Referenced Tests; Elementary Secondary Education; *Mathematics; *Multiple Regression Analysis; Norms; Predictor Variables; Program Evaluation; *Reading; School Demography; School Districts; *Schools; Standardized Tests; Test Results
IDENTIFIERS Arizona (Mesa); *Mesa Public Schools AZ

ABSTRACT

This paper addresses the use of multiple regression in the analysis of reading and math scores. The data presented in the paper were from the Mesa Public Schools in Arizona, and were compiled during the 1975-76 school year. Five criterion factors were selected for investigation: (1) third grade Stanford reading scores, (2) third grade criterion reading scores, (3) third grade criterion math scores, (4) fifth grade Stanford math scores, and (5) sixth grade criterion math scores. The criterion reading and math scores were for the Mesa-developed criterion reading and math tests. Other data was then summarized to form eleven demographic factors that could be used to predict the value of these criterion variables for each of the schools for the 1975-76 school year. Through the use of multiple regression, predicted or expected values were determined for individual schools. These values were then used to compare with the actual scores the school attained. Through the use of multiple regression an individualized predicted value or mean value was determined for each school based on its demographic characteristic. Comparisons of achievement of the school were then made based on its individual characteristics. This method was then contrasted with simple comparisons of a school's achievement with the district mean value and national norm mean. (Author/MV)



research and evaluation

A New Yardstick for Comparing Achievement in Basic Skills

JAMES S. DE GRACIE
FRANK L. VICINO

DR. GEORGE N. SMITH
SUPERINTENDENT
MESA PUBLIC SCHOOLS

DR. JAMES K. ZAHARIS
ASSOCIATE SUPERINTENDENT
EDUCATIONAL SERVICES
MESA PUBLIC SCHOOLS

PRESENTED AT THE:
1977 AERA ANNUAL MEETING
APRIL 1977

Foreword

The Mesa Public School District covers grades K through 12. It serves approximately 30,000 students with 25 elementary schools, 6 junior high schools and 4 senior high schools.

It is the major function of the Mesa Public Schools Department of Research and Evaluation to provide information to decision makers. This information should be timely, valid, reliable, and of sufficient depth, yet concise enough to speak squarely to the issues at hand. The decision makers noted above range from the classroom teacher to the superintendent of the district, depending upon the need for specific information. Persuant to the accomplishment of this task, the Department has developed what they consider a viable method of comparing achievement in basic skills at the unit level. This report then is the type of report that is furnished to district administrators so that they in turn can develop specific objectives at the unit level based on as much information as possible.

Introduction

One of the major problems in determining if a given basic skill program at a school is performing up to expectation is in determining what yardstick is available for measuring the achievement in the basic skill area. A number of yardsticks or methods are certainly available and used. One is the use of longitudinal data to determine program effectiveness. Although this type of information can be used to determine changes over years, assuming the populations have remained constant, the information cannot by itself be used to determine the effectiveness of a given program. One other method used is com-

paring the scores at a given school with a norm value. The drawbacks in this case are the same as those whenever norm-referenced data are used, that is, is the norm a realistic one for the particular district and the individual school that is being used in the comparison? As can be seen from Figure 1, when comparing the elementary schools in the Mesa district to the norm mean all but two schools equal or exceed the norm. This could lead to the conclusion, which may be erroneous, that everything is fine in River City. The students in the Mesa district, however, have above average I.Q. scores, the median income of the families are above average and the percent of minorities is lower than the national average. If these and other factors are incorporated in the analysis, one must question the advisability of resting on the laurels of the schools' reading scores. Another method to determine program effectiveness is to set or determine a given criterion for a particular school. This method also has drawbacks, the main ones being the difficulty of determining an individual school criterion and the lack many times of objectivity in the task of setting realistic objectives for determining program effectiveness at an individual school.

Still another method would be to use the district mean as the standard and compare all the schools' achievement with the mean score plus or minus some tolerance on this particular achievement factor. Figure 1 shows what happens when this method is used. The mean results of the third grade Stanford reading test are presented by school. The tolerance was determined by going three standard errors above and below the mean;

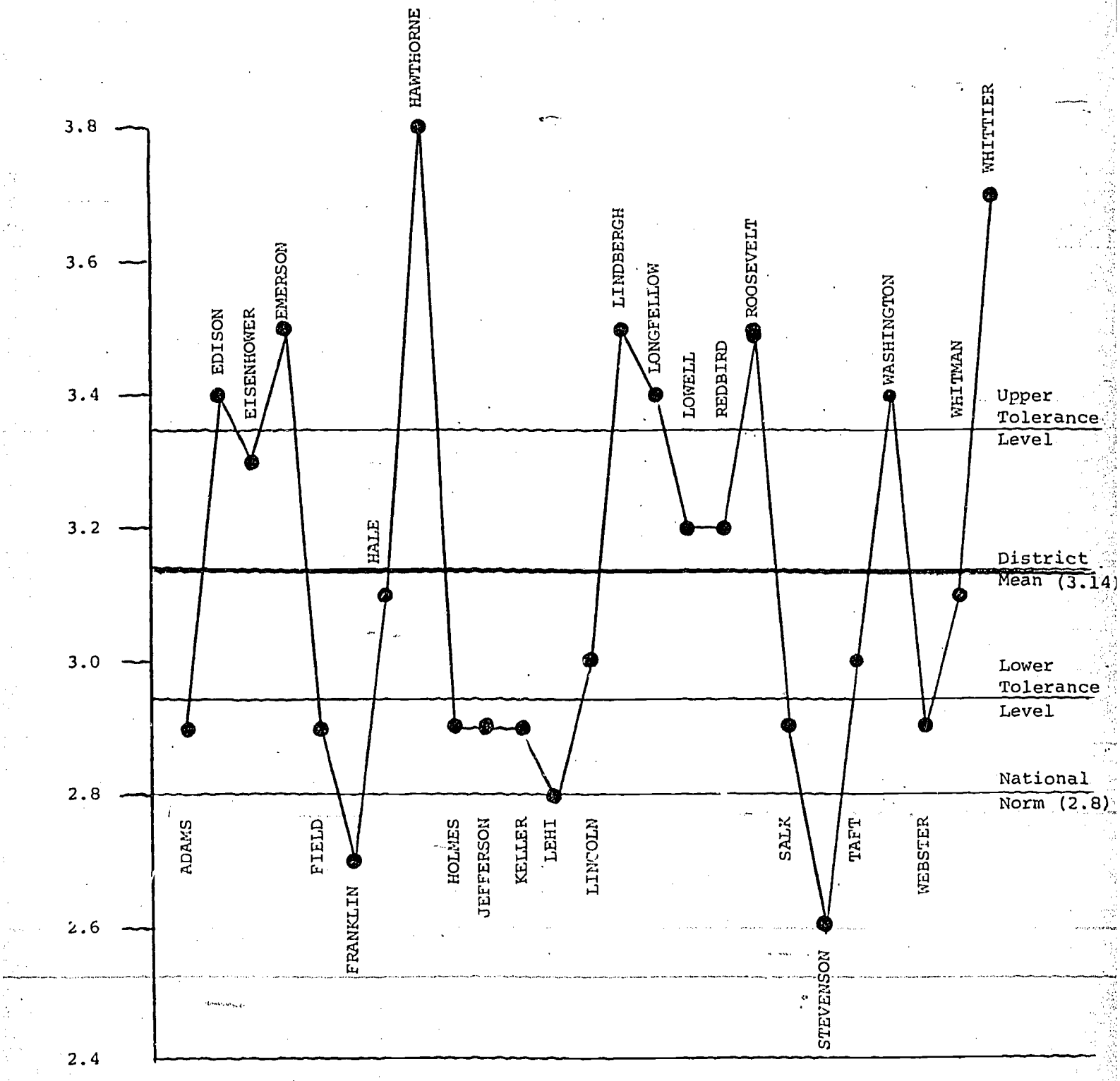


Figure 1

Mean Third Grade Stanford Reading Scores by School
with Corresponding Tolerance Interval About District Mean

that is, tolerance limits of 2.95 to 3.33 about a mean of 3.14. As can be seen, 7 of the 25 schools fall within the range about the mean with 8 above and 10 below the range. Thus, it would appear that student achievement in the Mesa district is at least satisfactory for 60% of the schools. Here again, as in the example where a national norm value was used, there is not sufficient information to judge that the reading program at Washington is outstanding, that at Hale satisfactory and that at Franklin poor. To really consider the programs at these schools one must consider that the mobility rate at Frankling is over 50% during the year, i.e., more than 50% of the students at Franklin are new to the school during the school year. The absence rate at Franklin is among the highest in the district. The converse of these factors is true for Washington with Hale somewhere in between. When one considers the reading scores compared to the district mean, therefore, is this a good yardstick for measuring program effectiveness?

It is not being said that the above examples of the use of data or information are not valid and useful methods of using the information. What is being said is that there is a need for a new yardstick for comparing achievement in the basic skill areas. One that attempts to overcome the deficiencies noted in the methods previously discussed. This is not to say that one method can overcome all deficiencies, but it is apparent that a need exists for an additional yardstick to fill the gap in the current measuring devices. This paper, then, addresses this need and proposes the use of multiple regression techniques in developing such a yardstick.

Method

One of the major problems in determining if a given program at a school is performing up to expectation is in finding comparable data on which to base the decision. In this document, an attempt is made to furnish the building administrators with data that can be used to make this kind of judgment.

Six criterion factors were selected for the investigation. These factors are: third grade Stanford reading scores, fifth grade Stanford reading scores, third grade criterion reading and math scores, and sixth grade criterion reading and math scores. Data were then collected on factors that could be used to predict the values of these factors for each school for the 1975-76 school year. The values of these factors can be found in the Mesa School Profile 1975-76. Step-wise multiple regression techniques were then used to determine which were the best prediction factors (see Appendix). These factors were then used to come up with a predicted value or an expected value for a given school. The prediction equations take into account the selected demographic factors or prediction factors in arriving at the predicted value. The building administrator can then use these predicted values to compare to the actual values for last year.

The definition of the different criterion and prediction factors that were used in this study are given below.

Criterion Factors

1.	Third Grade Stanford Reading	This is measured by the average third grade reading score achieved on the Stanford given in October 1975-76.
2.	Fifth Grade Stanford Mathematics	This is measured by the average fifth grade math score achieved on the Stanford given in October 1975-76.

3.	Third Grade Criterion Reading	This is measured by the average percent correct attained by the third grade students on the district criterion-referenced reading test administered in May.
4.	Third Grade Criterion Mathematics	This is measured by the average percent correct attained by the third grade students on the district criterion-referenced math test administered in May.
5.	Sixth Grade Criterion Reading	This is measured by the average percent correct attained by the sixth grade students on the district criterion-referenced reading test administered in May.
6.	Sixth Grade Criterion Mathematics	This is measured by the average percent correct attained by the sixth grade students on the district criterion-referenced math test administered in May.

Prediction Factors

1.	School Type	School type was simply coded by type of school. This factor was included more for prediction of plant vandalism then cognitive scores. The newer plant designs are more vandalism preventative.
2.	Student Inward Mobility Rate	This was measured by taking the aggregate total of the new students during the school year and dividing that by the aggregate total of all students at that specific school during that year.
3.	Ethnicity	This was measured by taking the total number of Anglo students at the end of the school year and dividing by the total number of students at the end of the school year.
4.	Student Intellectual Level	This is measured by the average second grade score on the Otis-Lennon test given in October 1975-76
5.	Expenditure per Student	This was measured by the total teacher expenditure plus the expenditure for books and supplies for a given school divided by the student population at that school.

6.	Student Absence Rate	This was measured by taking the total number of student absences at the school and dividing by the total student population at the given school.
7.	Teacher Academic Level	This was measured by taking the total teacher population at a given school at the end of the school year and dividing it into the total number of teachers with only bachelors degrees.
8.	Teacher Experience Level	This was measured by taking the total number of teachers with over 6 years of experience at the end of the school year and dividing by the total number of teachers at the given school.
9.	Teacher Ethnicity	This was measured by taking the total number of Anglo teachers at the end of the school year and dividing by the total number of teachers at the end of the school year.
10.	Teacher Absence Rate	This was measured by taking the total number of teacher days absent and dividing by the total number of teachers at the given school.
11.	Percent of Openness	This was measured by taking the estimated physical percent of openness at each school.

In addition to the predicted values for each school the standard error of these estimates were determined.

Results

Table I presents the actual values for each of the cognitive criterion factors along with the corresponding determined predicted values for the factors. Those cases in which the actual value fell below the predicted value by 3 times the standard error of the estimate are highlighted. In the case of the third grade criterion math, none of the prediction factors were found to have high predictability. The best estimate of an expected

Table I

Actual versus Predicted Values
for Selected Cognitive Criterion Factors

School	3rd Grade Crit. Math		3rd Grade Crit. Rdg.		Stan. 3rd Gr. Rdg.		Stan. 5th Gr. Math		6th Grade Crit. Math		6th Grade Crit. Rdg.	
	Act.	Pred.	Act.	Pred.	Act.	Pred.	Act.	Pred.	Act.	Pred.	Act.	Pred.
Adams	87	80	86	82	2.9	3.2	5.2	5.1	64	69	86	89
Edison	80	80	86	88	3.4	3.3	5.5	5.6	76	71	92	89
Eisenhower	69	80	71	79	3.3	3.2	5.4	5.2	66	69	86	87
Emerson	76	80	83	86	3.5	3.3	5.5	5.3	79	71	94	90
Field	82	80	84	82	2.9	3.2	5.3	5.1	67	68	88	89
Franklin	84	80	83	81	2.7	2.8	4.8	4.8	63	62	85	83
Hale	85	80	84	83	3.1	3.3	5.3	5.2	70	69	88	89
Hawthorne	87	80	89	85	3.8	3.4	5.6	5.6	70	72	90	89
Holmes	88	80	84	85	2.9	3.0	4.9	4.9	60	65	86	86
Jefferson	79	80	78	82	2.9	2.9	4.9	5.1	66	66	88	87
Keller	82	80	86	82	2.9	2.6	4.9	4.5	78	62	90	86
Lehi	70	80	79	85	2.8	3.0	4.5	4.7	59	68	87	89
Lincoln	70	80	80	82	3.0	3.0	4.8	5.0	66	66	89	87
Lindbergh	93	80	91	82	3.5	3.3	4.5	5.0	76	68	91	89
Longfellow	72	80	81	83	3.4	3.1	5.4	5.2	73	67	90	87
Lowell	84	80	81	80	3.2	3.2	5.1	5.2	68	71	89	90
Redbird	79	80	81	80	3.2	3.1	4.9	4.7	65	68	92	90
Roosevelt	84	80	88	87	3.5	3.5	5.8	5.7	79	74	95	91
Salk	80	80	80	80	2.9	3.0	4.7	4.9	63	67	87	88
Stevenson	66	80	70	77	2.6	2.7	4.7	4.6	52	62	80	85
Taft	84	80	81	76	3.0	3.0	4.5	4.8	68	65	87	87
Washington	82	80	86	83	3.4	3.5	5.5	5.6	72	72	87	90
Webster	78	80	78	83	2.9	3.3	5.4	5.3	67	69	84	88
Whitman	90	80	89	85	3.1	3.2	4.9	5.2	66	68	85	88
Whittier	73	80	85	86	3.7	3.4	5.8	5.4	65	70	86	89

value is, therefore, the district mean which is presented in Table I.

To illustrate this method further, Figure 2 presents a graphic example of the results. Third grade mean Stanford reading scores are presented by school. The expected or predicted values are also plotted with corresponding tolerance intervals. By comparing Figure 1 with Figure 2 the difference in the results of the use of different yardsticks is quite evident. For example, Hale School was reading significantly above the national norm and well within the tolerance limits about the district mean. When the demographic characteristics of the Hale students are taken into account, however, and a predicted value for the Stanford third grade reading scores was determined, it can be seen from Table I and Figure 2 that Hale School is in fact reading below their predicted value and significantly below. By contrast, Franklin School, which when using the two other yardsticks was reading below the national norm and significantly below the district mean, was within its individual tolerance level when the demographic characteristics of the students were taken into account through the use of a multiple regression framework. The third school, Washington, which was used in the previous example was reading, 34, far above the national norm of 2.8, significantly above the district mean of 3.1; but when the demographic characteristics of the students were incorporated in the analysis, it was found that the students at Washington were reading below what would be expected of them. The value was still within the individual tolerance level, but it certainly does not show the outstanding

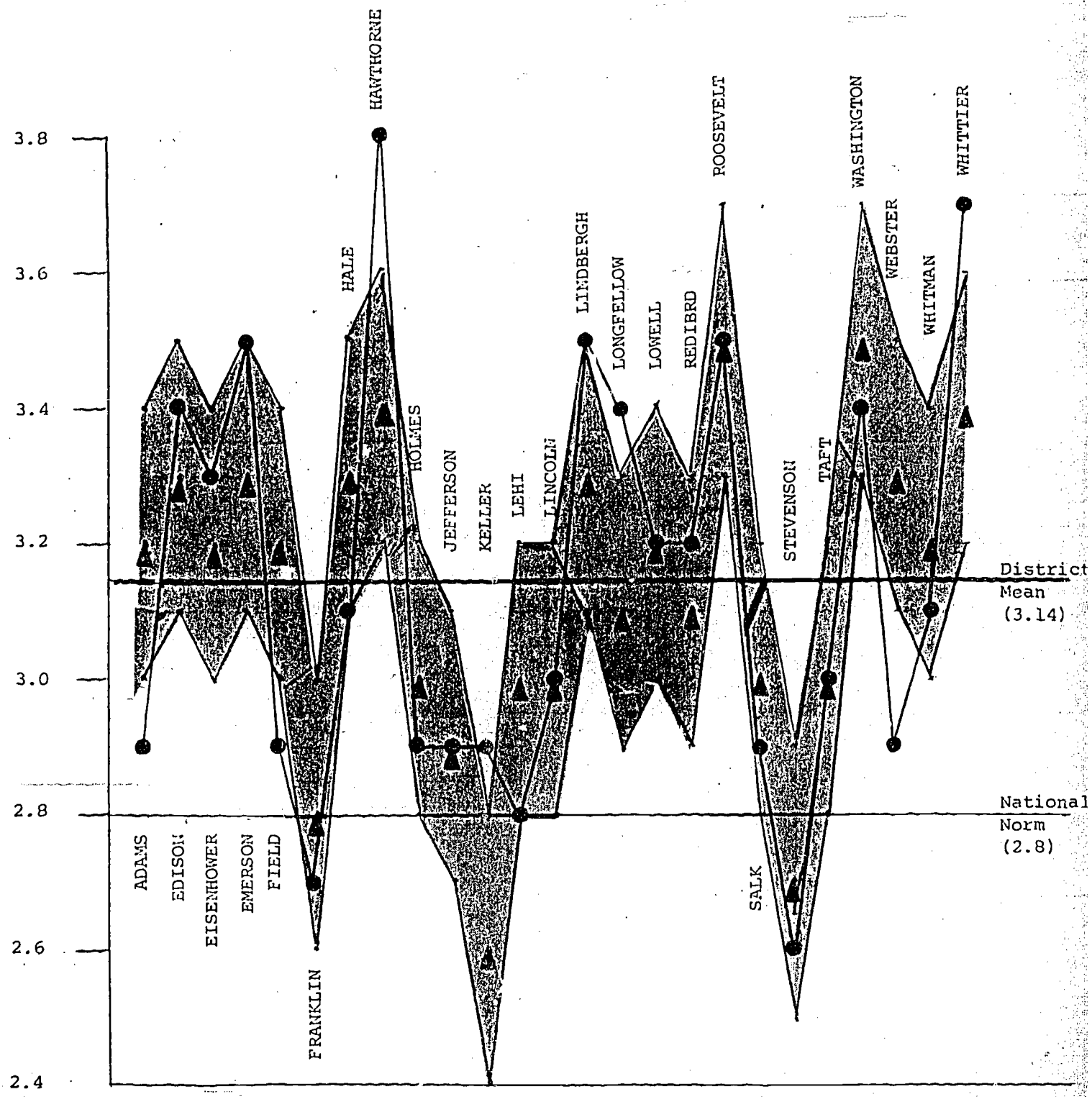


Figure 2

Mesa Third Grade Stanford Reading Scores by School with Corresponding Individual Tolerance Intervals

performance that was evidenced when compared to the national norm or district mean. Similar procedures and analyses are used for each of the criterion factors when analyzing the programs for each of the individual schools.

Carrying the analysis one step further, school achievement versus their potential as determined by the predicted values were summed over the six criterion factors. A simple plus minus scheme was used in this analysis. The deviations, being coded simply as a plus 1 or a minus 1, were summed over the six criterion variables and the schools were grouped according to:

Schools performing above potential

Schools performing at potential

Schools performing below potential

Table II presents the results of this analysis.

Summary and Conclusions

The Mesa Public Schools have been using this yardstick for comparing achievement in the basic skill areas for the last three years. It certainly is not used in isolation, but is one more added bit of information furnished to the superintendency so that they in turn can work with the unit level administrators in setting up specific goals and objectives at the unit level. Other measures are also furnished to the superintendency. The district mean is taken into account, the national norm value is considered, longitudinal data is examined, but when all is said and done the demographic characteristics of the students or the clientele that is actually served at the school must be taken into account when examining program effectiveness at

TABLE II

School Achievement versus Potential

Schools Performing Above Potential (Sum of deviation direction +6 to +2)	Edison Emerson Franklin Hale Hawthorn Keller Lindbergh Longfellow Redbird Roosevelt Taft
Schools Performing at Potential (Sum of deviation direction +2 to -1)	Adams Field Holmes Jefferson Lincoln Lowell Washington
Schools Performing Below Potential (Sum of deviation direction -2 to -7)	Eisenhower Salk Stevenson Webster Whitman Whittier Lehi

this level. It is felt that through the use of multiple regression this is done in a systematic manner, thereby insuring the preclusion of subjective bias.

The district feels that although this is a very useful and viable technique, it is certainly not the end. The multiple regression techniques used to determine the predicted values are based on school values [illegible] on individual student values. It is felt that a more vali[illegible]eliable measure will come with the use of individual student data. School predicted values will then be determined by weighting by the proportional number of students with those demographic characteristics at the individual school. Attempts are being made currently to generate a viable student data base which can be used in such a process. Another problem area is the use of grade equivalency scores in the analysis of the Stanford third grade reading and fifth grade mathematics. It is realized that raw scores or standard scores would be better measures to use in the analysis, but Mesa, as is the case in other districts, has felt the pressure of using a variable that the general public seems to accept. Strides are also being made in this area to re-education the public to the use of other scores and the misuse of the grade equivalency score.

ADDENDUM

Table III presents the rankings of the schools on each of the six cognitive criterion factors. The rankings were determined by finding the deviation of the actual scores from the predicted scores. In addition to the rankings given for each of the individual criterion factors, an overall ranking of the schools was determined by summing across the ranks of the individual criterion factors and reranking to determine which school performed the best as measured by the deviation of actual score from the predicted score over the six criterion factors. In all cases, a rank of one was given to that school with the best performance.

Table IV presents the overall rankings of the schools for the 1973-74, 1974-75 and 1975-76 school years.

TABLE III

School Rank on Achievement Factors

School	3rd Grade Crit. Math	3rd Grade Crit. Rdg.	6th Grade Crit. Math	6th Grade Crit. Rdg.	5th Grade Stan. Math	3rd Grade Stan. Rdg.	Over-all
Adams	4.5	4.5	22	21.5	10	23.5	17.5
Edison	14.5	18	5.5	4.5	17	8	9.5
Eisenhower	24	25	18	16	5	8	21
Emerson	19	20	2.5	2	5	5.5	5
Field	12	8.5	13	16	5	23.5	12
Franklin	8.5	8.5	8.5	7.5	14	17.5	7
Hale	6	11.5	8.5	16	10	21.5	11
Hawthorne	4.5	4.5	15	10.5	14	1	3.5
Holmes	3	15.5	22	12.5	14	17.5	16
Jefferson	16.5	21	11	10.5	20.5	12	19.5
Keller	12	4.5	1	2	1.5	3	1
Lehi	22.5	23	24	19	20.5	21.5	25
Lincoln	22.5	18	11	7.5	20.5	12	19.5
Lindbergh	1	1	2.5	7.5	25	5.5	2
Longfellow	21	18	4	4.5	5	3	6
Lowell	8.5	11.5	18	16	17	12	13
Redbird	16.5	11.5	18	7.5	5	8	8
Roosevelt	8.5	11.5	5.5	2	10	12	3.5
Salk	14.5	14	20	16	20.5	17.5	22
Stevenson	25	24	25	25	10	17.5	24
Taft	8.5	2	7	12.5	23.5	14	9.5
Washington	12	7	11	21.5	17	17.5	17.5
Webster	18	22	15	24	10	25	23
Whitman	2	4.5	15	21.5	23.5	17.5	15
Whittier	20	15.5	22	21.5	1.5	3	14

TABLE IV

Overall Rank on Achievement
Factors for 1973-74, 1974-75 and 1975-76

Schools	1973-74 Rank	1974-75 Rank	1975-76 Rank
Adams	3	3	17.5
Edison	9	11	9.5
Eisenhower	11.5	4	21
Emerson	20	22	5
Field	17	23	12
Franklin	16	6	7
Hale	4	5	11
Hawthorne	11.5	7	3.5
Holmes	21	17	16
Jefferson	8	12	19.5
Keller	--	--	1
Lehi	15	21	25
Lincoln	19	1.5	19.5
Lindbergh	23	20	2
Longfellow	1	9	6
Lowell	22	16	13
Redbird	7	15	8
Roosevelt	2	8	3.5
Salk	18	10	22
Stevenson	--	18	24
Taft	13.5	14	9.5
Washington	--	--	17.5
Webster	24	13	23
Whitman	13.5	1.5	15
Whittier	10	19	14

APPENDIX

FACTORS USED IN PREDICTIONS

Third Grade Criterion Reading Test	Multiple R (Accumulated)
Student Inward Mobility Rate	0.469
Percent of Teachers with BA	0.575

Sixth Grade Criterion Math Test	Multiple R (Accumulated)
Otis IQ	0.485

Sixth Grade Criterion Reading Test	Multiple R (Accumulated)
Otis IQ	0.493
Teacher Absence Rate	0.555

Fifth Grade Stanford Math Test	Multiple R (Accumulated)
Otis IQ	0.721
Teacher Absence Rate	0.790
Percent of Anglo Students	0.822

Third Grade Stanford Reading Test	Multiple R (Accumulated)
Otis IQ	0.729
Student Absence Rate	0.773