DOCUMENT RESUME

ED 312 302 TM 014 084

AUTHOR	Fetler, Mark
TITLE	A Method for the Construction of Differentiated School Norms.
PUB DATE	20 Oct 89
NOTE	18p.; Paper originally presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; Grade 8; Grade 12; *Local Norms; Performance; Public Schools; Reading Achievement; Reading Tests; School Demography; *School Effectiveness; Science Tests; Secondary Education; *Secondary School Students; *Socioeconomic Status; State Norms; Statistical Analysis
IDENTIFIERS	California Achievement Tests; *Differentiated School Norms

ABSTRACT

Empirical norms are a traditional and accepted tool for reporting performance. Other methods (e.g., statistical regression) are used more often in a researc' context to make conditional statements about performance given a certain demographic context. This paper illustrates a method for the construction of differentiated school norms and the computation of percentiles on the basis of demographic background factors. These differentiated norms facilitate reports of performance that take into account demographic conditions that are beyond the control of schools. With this approach each school is located at the median of its own norm group with regard to an index of socioeconomic status. The computation of percentiles is illustrated with science achievement test scores for all eighth-grade students in 1,573 public schools in California. The effects of varying the size of the norm group are illustrated with reading achievement test scores for all 12th-grade students in 797 public high schools in California. The computing and reporting of differentiated norms are complex and may be difficult to communicate to general audiences. Local authorities in California, however, have reported few difficulties so far in communicating the concept of differentiated norms. Two tables and one figure illustrate the study data. (Author/SLD)

A Method for the Construction of Differentiated School Norms

Mark Fetler

California State Department of Education

October 20, 1989

Opinions presented here are not necessarily those of the California State Department of Education. This paper was originally presented at the 1989 annual meeting of the American Educational Research Association in San Francisco.

Abstract

Empirical norms are a traditional and accepted tool for reporting performance. Other methods, e.g., statistical regression, are used more often in a research context to make conditional statements about performance given a certain demographic context. This paper illustrates a method for the construction of differentiated school norms and the computation of percentiles on the basis of demographic background factors. These differentiated norms facilitate reports of performance which take into account demographic conditions which are beyond the control of schools. With this approach each school is located at the median of its own norm group with regard to an SES index. The computation of percentiles is illustrated with science achievement test scores for a sample of 1,574 schools. The effects of varying the size of the norm group are illustrated with reading achievement test scores from a sample of 797 high schools.

A Method for the Construction of Differentiated School Norms

Norms permit an assessment of the performance of an individual or group of individuals relative to a defined population. The norms can represent either an entire population or some meaningful subgroup. Commonly, norms are intended to represent an entire population although some test publishers offer norms based on a subgroup, for example, local norms which are based on individuals or groups in a limited geographical area. According to Angoff (1984, p. 49), "the value of school mean norms can be enhanced if they are further differentiated in terms of school and community variables." Cronbach (1984, p. 111) notes that populations can be subdivided to allow differentiated comparisons of schools with reference groups of similar characteristics. An advantage of differentiated norms is that they can provide users with additional information which is helpful in understanding school performance. The objective of this study is to illustrate a method for constructing norms which take into account social and demographic characteristics of schools. Two procedures which have been traditionally used for making differentiated comparisons of schools are briefly sketched below, including expectancy tables and linear regression. This is followed by a discussion and illustration of differentiated norms using science achievement scores and a comparison with the use of science residual scores.

One simple method of subdividing a population of schools is to compute and use a socioeconomic index to sort schools into fixed subgroups or SES ranges. Groups are defined on the basis of their location in fixed ranges of the value of a predictor, here a measure of socioeconomic standing. For each of these subgroups the distribution of the criterion variable is used to make normative judgements about school performance, i.e., the computation of a percentile rank. This is the expectancy tables described by Angoff (1984). A disadvantage of expectancy tables is illustrated by the example of two schools that differ just enough on the SES index to be classified in different subgroups. The criterion performance of these two schools could be identical, yet their percentile ranks would differ, depending on the subgroup. If the SES measure is strongly correlated with the criterion, a change in subgroup can result in a large change in the percentile rank. This type of boundary effect is an issue when comparing the performance of schools in different SES subgroups.

Another approach for making differentiated comparisons employs regression techniques. yer, 1966; Dyer, Lynn and Patton, 1969; Klitgaard and Hall, 1973; Marco, 1974; Fetler and Carlson, 1985) For example, school mean achievement is predicted from socioeconomic or demographic information describing the pool of test takers. The difference between the predicted and the actual achievement, a residual or adjusted score, can be interpreted as a comparison with other statistically similar schools, and as the school's own effect on performance. Residual scores have long been used by educational researchers for making statistically controlled comparions of individual or group performance. But residuals have not so far replaced percentiles for communicating to the general pubiic and ›ch‹ l officials.

The expectancy table method can be elaborated to construct differentiated norms for individual schools. (California Department of Education, 1988) Based on each school's unique social and demographic characteristics, a norm group is defined which consists of all other schools within a specified range in the distribution of a socioeconomic index. Unlike the expectancy table method, the range floats, so that each school is at the median of the distribution of the socioeconomic index for its group. This group is smalle' and less diverse in terms of SES than the original population of schools. Percentiles are computed with reference to the differentiated norm group. Such percentiles provide a means for making normative judgements about schools which take into account social and demographic background factors. Direct comparisons with other schools in the group can be made. The construction of differentiated norms is illustrated below.

Data Source

The California Assessment Program annually administers a Science achievement test to all eighth grade students in 1573 public schools as a part of mandated statewide testing. A Reading achievement test is annually administered to all twelfth grade students in 797 public high schools, as well. School average achievement is expressed as a scaled score computed using Item Response Theory techniques (Bock, Mislevy and Woodson, 1982). Background information is obtained from a student survey administered during the assessment. Students indicate on a five point scale the educational level attained by his or her most educated parent: (1) not a high school graduate; (2) high school graduate; (3) some college; (4) college graduate; (5) advanced degreee. The parent education index (PEI) is the average of all student responses. Student mobility (MOB) was estimated by the percent of students newly enrolled in the district during the last year. English language fluency was estimated by the percent of test takers classified by teachers as limited English proficient (LEP) according to state mandated criteria. Finally, a measure of poverty, the percent of families receiving Aid to Families with Dependent Children (AFDC) was obtained for each school.

Analyses

The analytical steps described below include: the computation of a composite SES index; the use of the composite SES index to construct differentiated norms and science achievement percentiles; and the computation of science achievement residuals to be compared with the percentiles. The composite SES index for a school is estimated as the predicted score from a regression which the parent education index, student mobility, percent of limited English speakers, and AFDC are predictors, and the average of eighth grade reading and mathematics achievement is the criterion. Linear regression is one practical means of determining weights for the SES component measures in order to produce a composite index. Given that the predicted score can be interpreted as that part of performance attributable to, or explained by the SES components, the prediction can function itself as a measure of composite SES. Fetler and Carlson (1985) have shown that regression weights obtained in this way are stable both across years and academic subjects. This particular method of computing an SES composite was chosen in order to obtain an index which is strongly correlated with the achievement test scores to be normed.

Schools are ranked on the basis of the composite SES index. The differentiated norm group for a particular school includes both the 10 percent of schools with immediately higher composite SES indexes and the 10 percent with immediately lower indexes. The differentiated norm group for a school in the highest 10 percent of the SES distribution consists of the top 20 percent of schools. An analogous rule is used for the schools in the lowest 10 percent of the SES distribution. The middle 80 percent of schools stand at the median of the SES distribution for their differentiated norm groups, but do not necessarily fall at the median for other measures. Once the differentiated norm group has been specified, a percentile is computed for a performance measure, e.g., achievement, using conventional procedures. Similar procedures were applied to the high school Reading achievement scores, except that total group sizes of 10, 20 and 30 percent were used.

Linear regression was used to estimate the science residuals needed for comparison with the percentiles. The school average parent education index, student mobility, percent limited English speakers and AFDC were the predictors, and science achievement was the criterion. The following model was obtained: Predicted Science Achievement = 146.9 - 0.8(AFDC) + 41.6(PEI) + 22.2(MOB) - 0.4(LEP). All parameter estimates were significant ($p < .05$), with R-square equal to 0.5. The science residual equaled the actual minus the predicted score.

Results

Descriptive statistics and correlations are shown in Table 1 for the composite SES index, science achievement, residual and percentile. SES correlated more highly with the achievement test score than with the percentile. Boundary effects in the highest and lowest differentiated norm groups are responsible for part of the correlation between the percentile and the SES index. Because schools in these groups are not at the median of the SES distribution there is a positive correlation between the SES index and the percentile. The SES index within each differentiated group does vary, which could also account for part of this correlation. The correlation between SES and achievement is smaller in the extreme groups than in the population. For the 333 schools (20 percent) with top ranking composite SES indexes, $r = .48$, and for the lowest ranking 333 schools, $r = .34$. This difference results from a restriction in the range of SES in the differentiated group and diminished variance compared to the population.

TABLE 1: CORRELATIONS OF SCHOOL SCIENCE ACHIEVEMENT AND SES

	SES INDEX	ACHIEVE-MENT	RESID-UAL	PERCENT-ILE
MEAN	505.6	263.1	0.00	50.22
SD	69.25	42.52	28.93	28.49
N	1574	1573	1573	1573
SES INDEX	1.00	0.73	0.00	0.13
ACHIEVEMENT	0.73	1.00	0.68	0.72
RESIDUAL	0.00	0.68	1.00	0.90
PERCENTILE	0.13	0.72	0.90	1.00

Descriptive statistics and correlations are shown in Table 2 for ranks obtained with the 10, 20 and 30 percent group sizes. The magnitude of the correlations suggest that manipulation of the group size had little effect on the ranking of schools. Even with larger groups most schools are still located at the median on SES and apparently remain at the same relative position on achievement. The ranks obtained with larger groups did correlate more highly with SES, an undesirable property, than those obtained with smaller groups. By definition, with larger group size there are more schools at the extremes that cannot be at the SES median.

TABLE 2: CORRELATIONS OF RANKS OBTAINED WITH DIFFERENT GROUP SIZES

	10 % GROUP	20 % GROUP	30 % GROUP	SES RANK
MEAN	50.2	50.1	49.9	399
SD	28.6	28.3	27.9	230
N	797	797	797	797
10 % GROUP	1.00	0.98	0.94	0.06
20 % GROUP	0.98	1.00	0.98	0.14
30 % GROUP	0.94	0.98	1.00	0.24
SES RANK	0.06	0.14	0.24	1.00

A distribution of percentiles for an undifferentiated norm group is by definition rectangular. The empirical distribution of the percentiles computed here is itself rectangular with a computed skewness of 0.0 and kurtosis of -1.2. Examination of the cumulative frequency distribution of percentiles revealed 26 percent with values at or below 25, 50 percent at or below 50, and 75 percent at or below 75. Although this result appeared likely, given the relative complexity of the differentiated norm group approach, an empirical confirmation is of interest.

Percentiles are plotted against science residuals in Figure 1. The relationship between the science percentile and residual resembles an elongated "S", roughly linear in the midranges and curved in the extremes. The curving in the extremes is mainly attributable to the upper and lower bounds inherent in percentiles by contrast with the theoretically unbounded residuals. Examination of the plot suggests that the two methods produce results that are fairly consistent. There are however a number of visual outliers throughout. This does not necessarily argue for the superiority of one method over the other. The existence of such outliers suggests that for some schools varying judgments could be reached using the two analytical methods. Where practical decisions are being made, the existence of such differences may be grounds for closer inspection of the raw data.

Descriptive statistics and correlations for the three types of twelfth grade ranks are shown in Table 2. The magnitude of the correlations between these types of ranks suggests that changes in group size, at least for the range of sizes investigated here, do not affect the ranking of schools. The percentiles obtained with larger groups are more strongly correlated with SES than those obtained with smaller groups.

Discussion

Residual analysis and differentiated norms can both be used to make differentiated comparisons among schools. The differences in these two methods suggest that they best serve different ends and a decision to use either should depend on specific needs. The use of residual analysis implies a comparison with other statistically similar schools. Part of the usefulness of this approach lies in the variety of statistical procedures available for examining, manipulating, making probability statements about, and interpreting residuals. However, residual analysis does not typically include the computation of percentiles and the associated normative judgements. (See, for example, Draper and Smith, 1966) The use of differentiated norm groups involves a comparison with an empirical distribution of scores. The resulting percentile permits normative statements about performance which take background factors into account. The following discussion describes selected technical issues related to the use of differentiated norms, including: the treatment of very nigh and low SES schools; the procedures for constructing differentiated norm groups; determination of optimal group size; and interpretive difficulties. Finally, a question about the appropriate use of differentiated norms is raised.

The problem with very high and low SES schools is that unique differentiated groups for them cannot be constructed and these schools do not fall at the median of the SES distribution of their group. Because SES correlates positively with performance the high SES schools will on the average receive an "unfairly" high percentile and the low SES schools will receive an "unfairly" low percentile, compared to other schools in the middle of the SES distribution. This problem is eased by the restricted range of SES in these extreme groups and the associated smaller correlation between SES and performance.

One issue in the construction of differentiated norms is the computation cf an appropriate composite SES index. Here linear regression was used to produce a composite score that reflected non-school influences on achievement. One advantage of using this type of predicted score regression is that the SES index will correlate with the criterion as highly as is possible under a least squares model. If the goal of using a differentiated norm is to adjust for SES, higher correlations are desirable. Other methods of constructing a SES index can be imagined, e.g., a purely logical evaluation of importance, or perhaps cluster analytic approaches could be used.

Another issue relates to the determination of optimal group size. Smaller groups permit finer differentiation and more appropriate comparisons. Larger groups, on the other hand, should result in in more stable estimates of the percentiles, other factors remaining equal. Manipulating the group size here did not change the ranking of schools. One interpretation is that whatever the group size, most schools are still at the median of their groups in terms of SES and maintain their relative position on achievement. The higher correlation of the percentiles with SES for the larger groups is related to the problems at the extremes of the SES distribution. With larger group sizes, the numbers of schools at the extremes, which cannot be at the SES median, increase. The results suggest that the determination of an optimal group size requires a balancing of the need for stable estimates of percentiles with the need to have percentiles which correlate at a low level with SES. In practice, both stability and correlation with SES will depend on the empirical distributions and covariance of the available measures. This implies that some trial and error investigation of different group sizes should precede any application of differential norms.

Finally, The computing and reporting of differentiated norms is complex and will almost always require computer technology. This may be a problem when communicating results to audiences who are accustomed to relatively simple, undifferentiated norms tables and percentiles which do not take into account background factors. On the other hand, some commercial test publishers will produce local norms for their clients on request. Local norms permit geographically differentiated comparisons with the performance of individuals or groups in a specific region. It is possible that the concept of differentiation on the basis of socioeconomic variables will as easily communicated as is the concept of local norms. Some support for this hypothesis may be found in the use made of differentiated norms by the California Department of Education (1988). Local school officials have reported few difficulties in communicating the concept of differentiated norms as described in this paper.

The use of differentiated norms could motivate statements that a particular school's performance is below expectation or above expectation compared to other similar schools. For example, a school's residual score could be positive or negative, or its percentile might be greater or lesst than fifty. Once the precise criteria for making these statements are described, they have explanatory value. However, these explanatory statements are sometimes taken further to imply the value judgements that performance above expectation is good, and performance below expectation is bad. Such value judgements may neglect to take into account what is desirable for schools in absolute terms thereby encouraging an unmerited complacency or undeserved censure. For example, a low SES school could receive a percentile greater than 50 or a positive residual score, and still not meet minimally acceptable standards for achievement. Although the school average achievement could be greater than that of a majority of similar schools, it could still be that the educational needs of its students are not well met. Such instances underscore the importance of distinguishing the appropriate use of differentiated norms to explain performance from the value judgements which might be made in addition.

References

Angoff, W. (1984). Scales, Norms, and Equivalent Scores. Princeton, New Jersey: Educational Testing Service.

Bock, R., Mislevy, R., and Woodson, C. (1982). The next state in educational assessment. Educational Researcher, March, 4-11.

California State Department of Education. (1988). Performance Report for California Schools. Sacramento: Author.

Cronbach, L. (1984). Essentials of Psychological Testing. New York: Harper & Row. 112-114.

Draper, N. and Smith, H. (1966). Applied Regression Analysis. New York: Wiley, pp. 86-103.

Dyer, H. (1966). The Pennsylvania plan. Science Education. 50, 242-248.

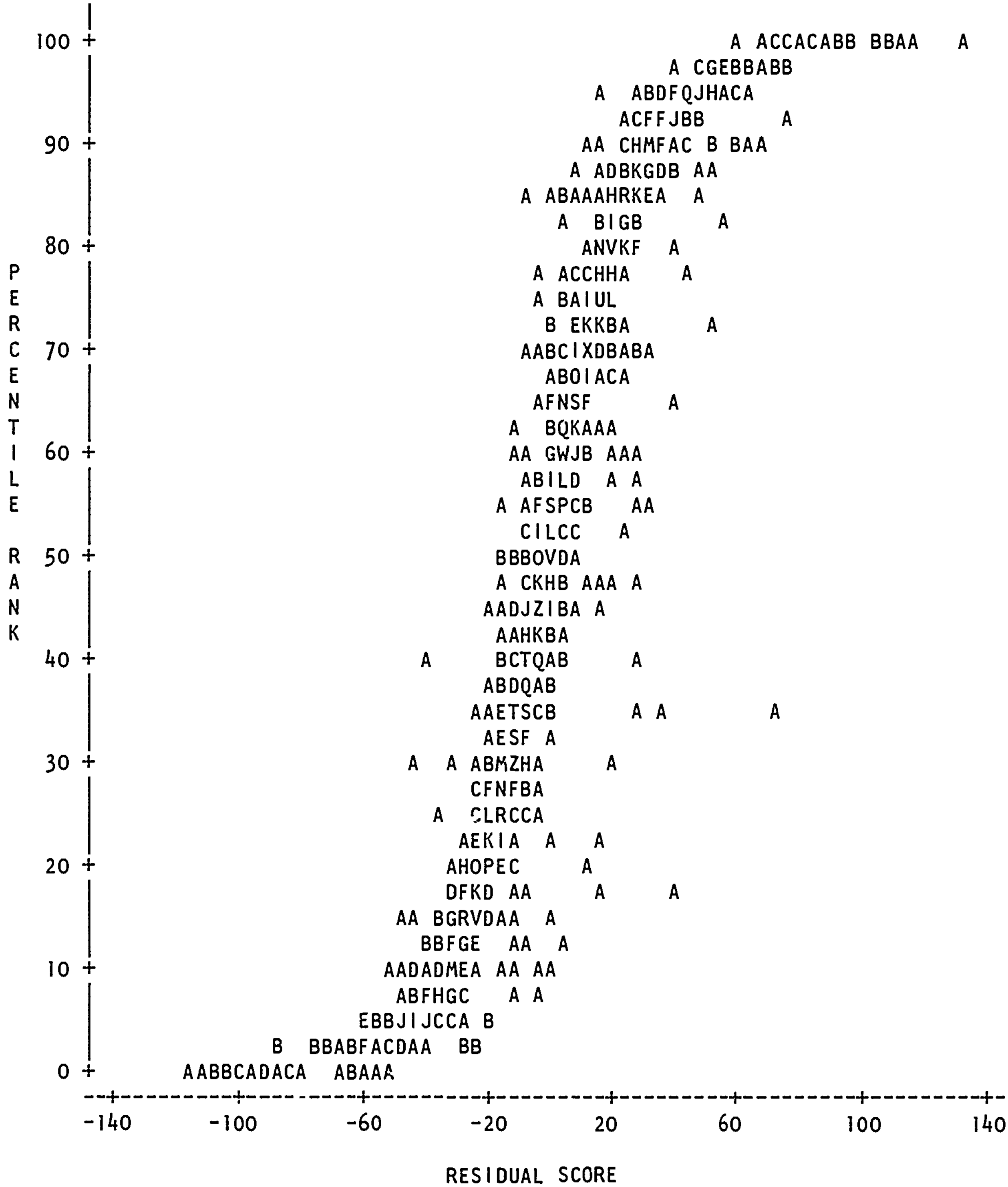
Dyer, H., Linn, R., and Patton, M. (1969). A comparison of four methods of obtaining discrepancy measures based on observed and predicted school system means on achievement tests. American Educational Research Journal, 6, 591-605.

Fetler, M. and Carlson, D. (1985). Identification of exemplary schools on a large scale. In Austin and Garber (Eds.) Research on Exemplary Schools. New York: Academic Press, 83-96.

Klitgaard, R., and Hall, G. (1973). A statistical search for unusually effective schools. Santa Monica, CA: Rand Corporation.

Marco, G. (1974) A comparison of selected school effectiveness measures based on longitudinal data. Journal of Educational Measurement, 11, 225-234.

FIGURE 1 PLOT OF SCIENCE ACHIEVEMENT PERCENTILE RANKS AND RESIDUAL SCORES



LEGEND: A = 1 OBSERVATION, B = 2 OBSERVATIONS, C = 3 OBSERVATIONS, ETC.