Articles/Research Reports

Decline in *ACT* Mean Scores: Alternative Explanations

John T. Mouw
John J. Cody
Gary L. Canivez

Abstract

The decline of standardized test scores has been for a number of years a major issue in the professional literature and the popular press. This study examines the feasibility that changes in demographic characteristics of the test population is a competing hypothesis to the claim that the quality of schools is decreasing. A sample of 2,000 respondents to the ACT in Illinois was drawn for each of the years 1971 through 1983. Those students who had responded to items regarding race, income, sex, and size of high school graduating class were eligible to be included in the sample. The findings included: (a) a greater proportion of students are taking the ACT; (b) the increase in test takers is from demographic groups which tend to score lower on standardized tests; (c) more than one-half of the sums of squares in the decline in test scores can be attributed to such demographic changes.

No report seems to have impacted the field of education in recent years as has the publication of *A Nation at Risk: The Imperative for Educational Reform* (National Commission on Excellence in Education, 1983). Numerous "indicators of risk" including statistics of illiteracy, underachievement of gifted students, lack of "higher order" thinking skills, and increased need for remedial mathematics in college were listed. However, the "indicator" that received the most attention by the media and in the professional literature was the decline in college admission test scores, especially scores on the *Scholastic Aptitude Test* (*SAT*).

The decline in *SAT* scores has been well documented by technical reports from the Educational Testing Service (College Entrance Examination Board, 1982). Similar declines in mean scores have been found

John T. Mouw and *John J. Cody* are faculty members in the Department of Educational Psychology at Southern Illinois University. *Gary L. Canivez* is a school psychologist in Phoenix, Arizona.

with the *American College Test* (Munday, 1976). In such studies, the mean performance of all tested students or the mean of certain demographically selected students was used as the index. However, such studies did not attempt to examine changes in mean scores in conjunction with the characteristics of the test-taking population.

A study by Powell and Steelman (1984) suggested that a simple observation of the mean performance might be misleading. Although they did not examine the decline of test scores over time, they did note limitations of using mean student performance as an accurate index. Their intention was to explain variability on the 1983 *SAT* performance across the 50 states using the following variables: (a) percent of students taking the test, (b) sex composition, (c) racial composition, (d) median income, and what they referred to as (e) "structural (school) variables." Their results showed that "82 percent of the difference among states appears to be an artifact of the proportion of students electing to take the *SAT*" (p. 400).

If variability among the states can be explained through differences in the characteristics of the tested populations, then predictable differences in the test scores across time should also be expected as the tested population changes. The result then should be a depressed mean if there is an increase in, for example, the proportion of low socioeconomic status (SES) test takers and if mean performance of low SES students is below that of other students.

Test performance does vary across a number of demographic variables. Low socioeconomic status populations obtain lower scores on achievement tests (Coleman et al., 1966; Jencks et al., 1972; Kagan, 1977; Mercy & Steelman, 1982). Black students tend to perform at lower levels than do their white counterparts (Crain & Mahard, 1978; Jensen, 1980). Males have been shown to perform at a higher level than females on tests of mathematical and spatial ability while the opposite has been observed on tests of verbal ability (Levine & Ornstein, 1983; Maccoby & Jacklin, 1974). School-related variables such as class size, academic course offerings, and school size (Russo & Checketts, 1978) have also been shown to be related to student achievement. If there is an increase in the proportion of black students, low SES students, or others with a confirmed history of lower performance taking the *SAT* or *ACT*, then a decline in mean scores should be expected. Just such an explanation was suggested by Jones (1981) and Fuerst and Hill (1981) and was supported by data presented by the Educational Testing Service (College Entrance Examination Board, 1977) using the mean scores of the test for various demographically defined subgroups.

Another way to examine the decline in mean scores requires an analysis in which pertinent demographic variables would be held constant or partialled out before examining variability of test performance

over time. Pertinent demographic variables would include traits such as those previously cited. Powell and Steelman's (1984) primary variable, "percentage of students taking the test," may be an umbrella that captures many of those known variables as well as some that are unknown. But such a variable suffers a bit in terms of explanatory value. Although the use of only known and available demographic variables to explain achievement patterns will err, it will likely be on the conservative side. Some potentially explainable variability will not be explained in the mean performance across time. For example, family income is an obtainable index of socioeconomic status, but suffers as a precise and reliable estimate of that package of education, books, library patronage, social participation, etc., that comprises *socio*economic status. Limitations of reliability lead to limitations of predictability. Therefore, the predictive relationship would tend to be underestimated.

A *truer* picture of declining test score means would entail an analysis of score changes independent of the aforementioned confounding variables. To that end, the purposes of this study are to (a) examine the results of student performance on the *ACT* in Illinois over a 13-year period, (b) control (through the use of statistics) selected demographic variables to determine their effects on the mean scores of the *ACT*, and (c) compare the mean scores adjusted for the effects of those demographic variables to the unadjusted mean scores. The results of such analyses will provide longitudinal data with which a more realistic description of mean score decline can be presented.

Method

To examine potential changes in demographic characteristics of the tested population and the effect any such change has on mean scores, a data set is required that spans a number of years and contains pertinent demographic information in addition to mean scores. Such a data set was obtained from the American College Testing program for the school years from 1971 to 1983 inclusive. The data tapes contained information on all tested students in Illinois for this period. In addition to the test scores for Mathematics, English, Natural Science, Social Science, and the Composite, each student record contained information from the student profile section.

The variables that were consistently available over the 13-year period and were potentially confounding variables included: (a) race, (b) parents' income, (c) sex, and (d) size of high school graduating class. Size of graduating class was coded into four categories: (a) 0 - 24 students, (b) 25 - 99 students, (c) 100 - 399 students, and (d) 400 or more students. Race was coded into five categories: (a) Afro-American (black), (b) American Indian, (c) Caucasian American (white), (d) Spanish-speaking American, and (e) Oriental or Asian American. Parents' income on the student profile section appeared as eight non-interval categories: the lowest was "less that $3,000" and the highest was "$20,000 and over." These categories, as recorded in the student profile, did not change to compensate for inflation until the 1981 school year. Therefore, the highest category in 1971 would not be financially equivalent to that same category in 1981. The categories were recoded in this analysis by prorating from the 1981 base year according to the Consumer Price Index (CPI). An income of $20,000 in 1981 was roughly equivalent to $15,000 in 1978 (the second category in the student profile), to $12,000 in 1974 (the third category), and to $9,000 in 1971 (the fourth category). To equate income level across the 13 years, the above values were used as minimum cutoffs for the *high* income level. Accordingly, the *middle* income level included these intervals: $6,000 - $9,000 (1971 - 1973); $9,000 - $12,000 (1974 - 1977); $12,000 - $15,000 (1978 - 1980); and $15,000 - $20,000 (1981 - 1983). The *low* income category included levels that were below the minimum cutoff of the *middle* income group.

Note that $20,000 in 1981 would not, normally, be the lower cutoff for a *high* income category. The three resulting categories could better be described as (a) very low income, (b) low income, and (c) middle to high income. Such necessary recoding resulted in a disappointingly imprecise index of economic, as well as socioeconomic, information. Such lack of precision leads to errors in measurement, i.e., unreliability. Because limited reliability results in limited predictability, the obtained relationship between the operational definition of income and performance is quite probably an underestimate of the true relationship between socioeconomic status and performance.

Two thousand student records were randomly selected from each of the 13 years. The two indices of race and parents' income are socially sensitive and prone to omission by students. Only records that contained complete information across the four demographic variables were retained. Almost 20% of the records contained no information regarding income; 7% omitted the item for race; 9% omitted both items. Those percentages remained relatively constant across the 13-year time period. The final sample consisted of 16,693 students, or approximately 1,284 students per year (see Table 1).

To confirm expectations regarding differences in performance among the various demographically defined categories, the mean score for the *ACT* Composite by category across the 13-year period was obtained. Second, an examination of the change in the proportion of students within the categories required three related variables. Data regarding the number of high school students in Illinois for each year were compared with the number of students who took the test (number of records on the annual tape) and the number of students within each of the categories for each of the years.

Table 1

Changes in the Proportion of Illinois Students Taking the ACT

Class of	No. of Students in School	No. of Students Tested	Estimated Proportion of Tested Students
1971	161,361	94,327	.585
1972	159,154	92,214	.579
1973	156,904	91,142	.581
1974	161,498	92,086	.570
1975	163,465	94,428	.578
1976	164,239	98,759	.601
1977	162,226	104,375	.643
1978	160,669	104,447	.650
1979	154,716	103,370	.668
1980	156,981	110,189	.701
1981	156,820	112,948	.720
1982	148,992	105,712	.710
1983	147,234	106,520	.722

The major research question was addressed by constructing general linear models in which the *ACT* Composite mean score was the dependent variable and the four demographic variables (in addition to the year) were included as predictor variables. All predictors including (a) year, (b) race, (c) parents' income level, (d) sex, and (e) size of high school graduating class were nominalized, i.e., used as class variables in the analysis. To capture the pattern of *decline*, a linear trend across the 13 years was obtained. In addition, a quadratic trend was obtained to examine the degree of recovery from the decline in later years.

Overall differences across the years as well as the linear and quadratic trends were examined in two ways. First, raw means and sums of squares among years were obtained to measure the decline or change as it is usually examined. Second, the four demographic variables were partialled out of the variable year. This yielded adjusted means and sums of squares indicating the change as though, hypothetically, the tested population had remained consistent across the years with respect to the four demographic variables.

Results

The number of students graduated from Illinois high schools fluctuated between 1971 and 1983. The largest number of students graduated in 1976 (164,239). The smallest number of students graduated in 1983 (147,234). However, the proportion of the population of high school students tested over these same years increased. In 1971, 94,327 high school students took the *ACT* as opposed to 106,520 in 1983. These data, presented in Table 1, clearly indicate an increase in the number of students taking the *ACT*. The change in the proportion of high school students taking the *ACT* in Illinois increased from approximately 58% during the years 1971 to 1975 to more than 70% during the years 1980 to 1983.

Not only did the proportion of test takers increase, but the demographic makeup of those tested also changed. While about 42% of the students were female during the years 1971 to 1973, Table 2 indicates an increase to as high as 54% in 1980, 1981, and 1982. While approximately 13% of the tested students were non-white in 1971, by 1983 the minorities comprised 20% of the group.

In spite of the ambiguity in the variable of parents' income, there was an identifiable change in the income levels of the tested population.

Table 2

Changes in the Percentage of Demographic Characteristics for the Tested Illinois Students, 1971 - 1983

Year	*N*	Sex *Female*	Income *Med-Hi*	Race *Non-white*	School Size *Large*
1971	1,186	41.9	60.4	12.8	45.4
1972	1,248	44.0	68.9	14.2	50.6
1973	1,291	41.8	72.1	13.2	49.8
1974	1,258	48.5	61.8	15.4	49.5
1975	1,230	52.4	65.6	13.5	49.5
1976	1,281	52.6	65.0	16.1	49.5
1977	1,340	55.8	69.0	15.8	46.9
1978	1,296	53.4	60.4	19.0	46.9
1979	1,342	53.0	65.7	17.9	47.3
1980	1,277	54.0	66.3	18.1	45.6
1981	1,272	54.4	57.5	16.0	43.9
1982	1,308	54.5	60.1	19.1	41.5
1983	1,364	52.5	57.8	20.0	40.3

Between the years 1971 to 1973 an average of 67% of the tested students came from the high income category; less than 60% were in that category from 1981 to 1983. (Recall that the lower limit of the *high* income category for 1981 to 1983 was only $20,000.) While almost half the students tested were from high schools with large graduating classes between 1971 and 1976, that category supplied only 40% of the tested students in 1983. This 10% change occurred despite the continued movement toward school consolidation in Illinois.

Changes in the demographic makeup of the tested population should be reflected in a change in the mean scores across years if there are differences among the means of the identified demographic categories. Table 3 presents the means across the four demographic variables used in this study. For each of the four variables, statistically significant differences were found between or among the means.

Table 3

ACT Composite Mean Scores for Various Demographic Categories for Students in Illinois, 1971 - 1983

Variable	Category	*n*	*M*	*F*
Race	Black	1,908	12.54	619.42
	Native American	105	16.69	
	White	13,943	20.29	
	Spanish American	483	14.48	
	Oriental American	254	18.55	
Sex	Female	8,546	18.50	176.40
	Male	8,147	18.94	
Income	Medium—High	10,504	20.51	271.67
	Low—Medium	3,272	18.27	
	Low	1,917	15.55	
Size of Graduating Class	<25	155	17.06	232.40
	25-99	1,842	18.66	
	100-399	6,954	18.60	
	≥ 400	7,742	19.92	

Male students performed better than females on the *ACT* Composite; white students had a higher mean score than did each category of non-white students; there was a linear increase in mean score performance as income increased; students from schools with large graduating classes performed better than students from schools with small graduating classes. The mean scores of students in the two middle categories of size of graduating class were between the extremes for that variable.

Given the variability in performance across categories of the four demographic variables and given the change in the proportion of students who comprise the annual tested population within those categories, one should expect a corresponding change in the mean score performance of the population across time. The following analysis examines that relationship.

First, raw mean scores on the *ACT* Composite were obtained across the 13-year period. Second, the least square or adjusted mean scores were obtained across the years. Such adjusted mean scores are estimates based on the assumption that demography is held constant (equal proportions) across the 13 years. The results appear in Table 4.

The raw mean scores began to drop from approximately 20 in 1974, decreased to a low of 18.64 in 1980, and then appeared to stabilize.

Table 4

Performance of Illinois Students on ACT, 1971 - 1983

Year	Raw *M*	Adjusted *M*	Diff.	Smoothed Raw *M*	Smoothed Adjusted *M*	Diff.
1971	20.38	20.09	.29	20.70	20.34	.36
1972	20.29	20.00	.29	20.28	20.00	.28
1973	20.50	20.05	.45	19.92	19.70	.22
1974	19.66	19.69	-.03	19.60	19.45	.15
1975	19.38	19.23	.15	19.33	19.24	.09
1976	18.78	18.77	.01	19.11	19.08	.03
1977	18.87	18.82	.05	18.94	18.93	.01
1978	18.72	18.95	-.23	18.82	18.89	-.07
1979	18.72	18.81	-.09	18.74	18.87	-.13
1980	18.64	18.77	-.13	18.72	18.89	-.17
1981	18.91	19.07	-.16	18.74	18.95	-.21
1982	19.06	19.36	-.30	18.81	19.06	-.25
1983	18.75	19.03	-.28	18.93	19.22	-.29

Although the adjusted mean scores followed a similar pattern, the discrepancy is not as large. Furthermore, the adjusted mean scores hint at a recovery from 1980 to 1983. The differences between the raw mean scores and the adjusted mean scores showed (a) the largest discrepancy to be in 1973, with a difference of .45; and (b) the expected tendency for the raw mean scores to overestimate the performance in the early years and underestimate the performance in the later years.[1]

The *decline in mean scores* implies a negative linear trend, while *stabilizing* implies a secondary positive quadratic trend. To examine these trends a model was run on both the raw mean scores and the adjusted mean scores across the 13 years. The results of this analysis yielded a smooth curve over the obtained mean scores. The trend over the unadjusted mean scores began at a high of 20.7 in 1971, decreased to approximately 18.75 during 1978 to 1980, and increased slightly to 18.9 by 1983. Assuming the demographic makeup of the tested population on the four selected variables was constant across the years, the trend would be estimated as beginning at 20.3 in 1971, decreasing to 18.9 during 1978 to 1980, and improving to 19.2 by 1983. According to the differences, the raw means were overestimates for our *average population* by .36 in 1971 while in 1983 the performance was underestimated by .29. Accordingly, .65 (.36 - (-.29)) of the 1.77-point (20.70 - 18.93) drop from 1971 to 1983 is explained by the four demographic variables.

An analysis of the variability among the means appears in Table 5 which contains the sums of squares for the various sources. A general

Table 5

Sums of Squares of ACT for Raw Scores and Adjusted Scores

Source	*df*	Raw Scores *SS*	Adjusted Scores *SS*	% Reduction
Year	12	6631.20	3559.01	46.33
Linear Trend	1	4285.38	1767.25	58.76
Quad. Trend	1	1686.82	1241.74	23.69
Deviation from Trends	10	659.00	550.02	16.54

linear model yielded (a) the sums of squares for differences among years, (b) the sums of squares of the raw means accounted for by the expected linear and quadratic trends, (c) the sums of squares among means after partialling out the effects of demography, and (d) the sums of squares of the expected trends after partialling out the effects of demography.

Variability among the means over the 13 years yielded a raw sums of squares of 6631.2. This was reduced to 3559.01 by adjusting for the four demographic variables indicating a reduction of 46.3%. The linear decrease over the 13 years in raw scores represented an *SS* equal to 4285.38, but this linear trend was reduced by 58.8% to an *SS* equal to 1767.25 when the demographic change in the population was considered. Finally, a slightly smaller recovery (quadratic trend) was found with the adjusted means (probably because there was less to recover). Little variability across either the raw or adjusted means other than the primary trends was found among the means.

Discussion

What was postulated by Fuerst and Hill (1981) and presented by the Educational Testing Service (College Entrance Examination Board, 1977) is empirically supported in this study. The demographic variables of sex, race, parents' income level, and size of high school graduating class explain almost one-half (46.3%) of the variability of the mean scores of the *ACT* Composite for Illinois students between 1971 and 1983. More importantly, the sums of squares associated with the decline (linear trend) of those means is reduced by 58.8% when those demographic variables are partialled from that pattern.

Race and sex are obvious demographic candidates for predictors. However, income level as reported on the student profile and recoded to account for inflation is an imprecise measure of important social characteristics such as parent education level or occupation. Size of high school graduating class may not be as useful a *school* variable as number of courses available or tax rate of the district. The argument might be made that the use of more salient and precise demographic predictors would explain even more of the decline in test score means. Such a conjecture seems plausible, although empirical support for it is not present in this study. What is apparent is that the population of tested students has changed over time and that the demographic makeup of the students can explain a large part of the decline in test scores.

The results of this study cast doubt on the validity of the claim that schools have become less effective. A descriptive analysis of the change in mean scores of the *SAT* or *ACT* is, in Campbell and Stanley's (1963) terms, a one-shot case study. Any causal interpretation, however implicit, is replete with competing hypotheses. The present study has isolated only some of those alternative explanations.

The *SAT* and *ACT* were not designed as instruments of school evaluation. Their intent was to assess a cumulatively acquired set of academic expertise that would yield a stable predictor of future academic performance for the individual student. The design of the instruments, the method of administration, and the subject pool of test takers are consistent for that purpose. If either the Educational Testing Service or the American College Testing program were assigned the task of evaluating the schools, the tests and the procedures would probably be quite different. No doubt a major change would entail the control of the sampling of subjects from all socioeconomic and socioethnic strata.

There are at least two alternatives for the solution to the problem of decreasing college entrance test scores suggested by this study. The first of these is to discourage students with mediocre academic achievement in high school from attempting the examination. Then, the population taking the test would probably be more similar to the early samples to which today's mean scores are compared. This study suggests that a decline, if present, would be minimal. Of course, such a solution would be absurd in the framework of the tradition of free access to public education that is practiced in America.

The second alternative is more deliberate and developmental. The schools could continue to serve as societal change agents according to the social and political mandates of the 1970s. Their curricula and services could be expanded to encourage the dream of equal educational opportunity for all. They would encourage all students to continue in high school and attempt to raise their achievement levels as well as their aspirations to attend college. However, necessary changes would require time. The reasonable solution to declining test scores in the face of the changing population requires education and patience while we work to change the academic expectations and performance of the groups who, for lack of academic and cultural exposure, tend not to score as high on standardized tests.

It is not the goal of this research to promote either of these alternatives. But, whatever the goal of the schools, the evaluation of their effectiveness must rest solidly on their purpose. What has been identified in this study is the fact that mean scores considered without other available data have been misleading. When students from all levels of academic potential, readiness, and knowledge are included in the *test* population, a likely result will be a depression in the performance of that expanded population. Mean scores should be expected to drop. Any attempt to enhance achievement must consider the present high school population that the schools are mandated to serve. And any attempt to assess the effectiveness of the schools must consider the same.

References

Campbell, D. T., & Stanley, J. C. (1963). **Experimental and quasi-experimental designs for research.** New York: Rand McNally.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J. M., Mord, A., Weinfield, F., & York, R. L. (1966). **Equality of educational opportunity.** Washington, DC: U. S. Government Printing Office.

College Entrance Examination Board. (1982). **National report on college bound seniors.** Princeton, NJ: ETS.

Crain, R. L., & Mahard, R. E. (1978). Desegregation and black achievement: A review of the research. **Law and Contemporary Problems, 42**(3), 17-56.

Fuerst, W. L., & Hill, R. R. (1981). Are *SAT* scores really declining? **Educational Technology, 21**(5), 43-45.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heynes, B., & Michelson, S. (1972). **Inequality.** New York: Basic Books.

Jensen, A. R. (1980). **Bias in mental testing.** New York: Free Press.

Jones, L. (1981). Achievement test scores in mathematics and science. **Science, 212**, 412-416.

Kagan, J. (1977). The child in the family. **Daedalus, 106**, 35-56.

Levine, D. U., & Ornstein, A. C. (1983). Sex differences in ability and achievement. **Journal of Research and Development in Education, 16**(2), 66-72.

Maccoby, E. E., & Jacklin, C. N. (1974). **The psychology of sex differences.** Stanford, CA: Stanford University Press.

Mercy, J. A., & Steelman, L. C. (1982). Familial influence on the intellectual attainment of children. **American Sociological Review, 47**(4), 532-542.

Munday, L. A. (1976). **Declining admissions test scores.** (ACT Research Report No. 71). Iowa City, IA: The American College Testing Program.

National Commission on Excellence in Education. (1983). **A nation at risk: The imperative for educational reform.** Washington, DC: U. S. Government Printing Office.

Powell, B., & Steelman, L. C. (1984). Variations in state *SAT* performance. Meaningful or misleading? **Harvard Educational Review, 54**, 389-411.

Russo, T. J., & Checketts, K. T. (1978). Three sets of ordered variables and their association to the *American College Test* scores. **Journal of Educational Research, 71**, 198-202.

Note

[1]Adjusted means yield an unbiased estimate of mean performance *assuming* (i.e., statistically) an equal proportion across years for the demographic variables of race, parents' income, size of high school graduating class, and sex.