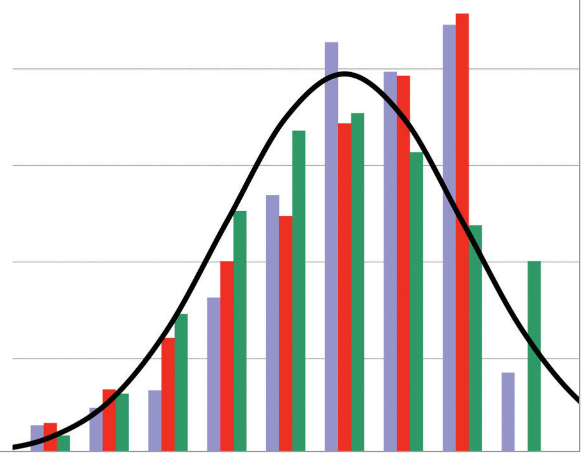
December 2013



ACT Research Report Series

2013 (9)

Interpreting Changes Over Time in High School Average ACT® College Readiness Assessment Composite Scores and ACT College Readiness Benchmark Attainment Rates

Richard Sawyer

For additional copies:
ACT Research Report Series
P.O. Box 168
Iowa City, IA 52243-0168

Interpreting Changes over Time in High School Average ACT® College Readiness Assessment Composite Scores and ACT College Readiness Benchmark Attainment Rates

Richard Sawyer

Abstract

Education officials and journalists frequently track changes over time in the average ACT® College Readiness Assessment Composite scores and ACT College Readiness Benchmark attainment rates of individual high schools. Using standard statistical methods, I examined how often changes in these statistics are unambiguously positive or negative, rather than plausibly due to chance (random variation). I studied two-year differences, five-year trends, ten-year trends, and the difference between the most recent five-year period and the preceding five-year period.

For a large majority of high schools, changes over the time periods studied were ambiguous: They could plausibly be attributed to random variation among student cohorts. For example, two-year differences in the average ACT Composite score were plausibly due to chance at 91% of schools; five-year trends were plausibly due to chance at 79% of schools; and ten-year trends were plausibly due to chance at 64% of schools. This result is also true of changes adjusted for student background characteristics and prior achievement.

As one would expect, unambiguous changes tend to be large and based on large numbers of ACT-tested students. This report describes simple ways for school officials to predict whether observed changes are unambiguous without doing a formal statistical analysis.

Acknowledgements

I would like to thank several people for their generous help in preparing this report. Kristy Watson and Edward Colby provided research on media stories about ACT scores. Paul Weeks and Lauri Chapel provided helpful information on statewide testing. Karen Zimmerman helped check the accuracy of transcriptions of the results. Jeff Allen, Valorie Foy, Julie Noble, Jim Sconing, Karen Zimmerman, and Robert Ziomek read previous drafts of this report and made many helpful suggestions for improving it

.

Interpreting Changes over Time in High School Average ACT® College Readiness Assessment Composite Scores and ACT College Readiness Benchmark Attainment Rates

Every year, ACT releases to individual high schools summary reports of their students' performance on the ACT® College Readiness Assessment (ACT, 2012a). Among the statistics contained in the reports are average scores in English, Mathematics, Reading, and Science, as well as an average Composite score. The reports also show the percentage of students whose scores indicate readiness to take credit-bearing first-year courses at typical postsecondary institutions, as well as information about students' background characteristics, high school course work, interests, and education plans. The College Board (2012) releases similar summary reports on the SAT.

Although the summary reports contain information on many student characteristics, education officials and local media pay particular attention to average test scores. Local media frequently compare the average scores of local schools, the state, and the U.S. According to ACT's communications staff, there were over 3,200 stories published in newspapers and magazines, on websites, or broadcast on radio and television, over the last five years about the average ACT Composite scores of local high schools. Media also pay close attention to *changes* in average scores and seek explanations for possible causes. A change of as little as 0.1 can be a cause for comment.

Year-to-year comparisons are based on different cohorts of students. Because comparisons of different cohorts are potentially influenced by concurrent changes in the cohorts' background characteristics and prior achievement, they are more difficult to interpret than "growth-model" comparisons that track changes over time in the achievement of individual students. For this and other reasons, year-to-year changes in the average scores at individual high schools are not *prima facie* indicators of changes in the schools' effectiveness. To encourage a longer-term perspective, the summary reports from both ACT and the College Board contain averages and percentages from the last five years. Nevertheless, most local public comment pertains to year-to-year changes.

Accountability Systems

A related use of test score summary data pertains to changes in proficiency rates (the proportion of students whose scores exceed a particular threshold). This use is related to the No Child Left Behind (NCLB) law, which requires states that receive Title I money from the U.S. federal government to develop performance standards in different skill areas for grades 3-8 and 11 in their public schools, to assess students' attainment of these standards, and to impose graduated levels of sanctions against schools whose proficiency rates do not demonstrate "adequate yearly progress" (AYP). As of 2013, two states used ACT test scores as part of their NCLB accountability system.

Under AYP, individual schools are expected to achieve targeted proficiency rates on their state assessments, based on the difference between their initial rate of proficiency and 100% proficiency, by 2014. Thus, AYP is a set of interim status goals, rather than a goal for yearly change (Twing, 2013). States were allowed to set their interim status goals in different ways, provided that they achieved 100% proficiency by 2014. Moreover, states were allowed great flexibility in using confidence intervals to determine whether individual schools demonstrate AYP (Davidson, Reback, Rockoff, & Schwartz, 2013). In recent years, however, these complexities to AYP have become moot in most states, because the U.S. Department of Education has granted waivers in exchange for states' efforts to improve schools and teacher effectiveness. As of August 2013, forty-one states had been granted waivers (McNeil, 2013).

Aside from NCLB, some states report changes in average test scores as part of their own accountability systems. The Nebraska Department of Education (2012), for example, reports performance indicators by school on a variety of measures. Among the indicators reported is the year-to-year change in average scores on Nebraska's state assessments. Although the department does not rank schools on their year-to-year changes by grade level (e.g., grade 6), it does rank them on year-to-year changes by grade configuration (e.g., grades 6-8). The department also compares the year-to-year changes of individual schools to those of the state as a whole.

College Readiness Benchmarks

Both ACT and the College Board also report to high schools the percentage of their students whose scores exceed certain thresholds indicating readiness to take typical first-year courses (ACT, 2012a; College Board, 2012). The ACT College Readiness Benchmarks (CRBs) are scores on the four component ACT tests (English, Mathematics, Reading, and Science) associated with a 50% probability of earning a B or higher grade in related credit-bearing first-year courses at typical postsecondary institutions (Allen & Sconing, 2005).[1] The SAT College and Career Readiness Benchmark is the SAT composite score (Critical Reading + Writing + Mathematics) associated with a 65% chance of earning a 2.67 or higher first-year GPA (Wyatt, Kobrin, Wiley, Camara, & Proestler, 2011). The SAT also has readiness benchmarks in each of its three content areas. In their summary reports to high schools, both ACT and the College Board report the percentage of students whose scores meet the benchmarks.

Norms for Changes in Average Scores

Ziomek (2000) calculated norms for the year-to-year change in high schools' average ACT Composite score, using data from the 1999 - 2000 graduating classes. The norms show that

[1] For recently updated values of the CRBs in Reading and Science, see Allen (2013).

although some schools had positive changes and others had negative changes, there was very little change (0.0 to 0.1) at typical high schools.

Statistical Precision of Changes in Average Scores

Media coverage of changes in average scores and proficiency rates does not often consider their statistical precision: Could year-to-year changes be plausibly explained by chance? If an observed change can be plausibly explained by random variation among student cohorts, it is difficult to argue that current students have learned more or less than students in previous years.

Variation in average scores or CRB attainment rates can be attributed to random measurement error in ACT test scores, to random variation in the students tested, and to systematic variation in the students tested. Random measurement error results from chance variation in how individuals respond to different test items or forms. Random variation in the students tested produces minor fluctuations in average scores or attainment rates over repeated sampling of students, but does not result from a change in the average level of students' achievement. A systematic change in the average level of students' achievement does not fluctuate over repeated sampling of students.

The psychometric characteristics of the ACT imply that random measurement error has a minor role in average score changes. The variability of the mean ACT Composite score for an individual high school depends on the within-school variance of the Composite score, which is typically about 17.6. The average standard error of measurement for the Composite score is .94 (ACT, 2007). Therefore, the average measurement error variance is approximately $.94^2 = .9$, and the proportion of variance within high schools that is associated with measurement error is approximately $.9/17.6 = .05$. Thus, most of the within-school variance of the ACT Composite score (and, therefore, most of the variability in its average) is due to change in the students tested, and not to random measurement error.

A principal goal of this paper is to determine, for a large representative sample of schools, whether the observed changes over time in their average ACT Composite scores and CRB attainment rates can plausibly be associated with systematic, rather than random, change in the students tested. A standard statistical tool for answering this question is to determine whether a 95% confidence interval about an observed change includes the value 0. If the 95% confidence interval includes 0, then in hypothetical repeated sampling of students, leaving everything else the same, the observed change could plausibly reverse sign (e.g., an increase in average score would become a decrease). It is difficult to interpret the meaning of a change that could plausibly be either positive or negative, other than to say that it might be small. If the 95% confidence interval about an observed change does not include 0, then I will call the observed change "unambiguous." More commonly, unambiguous changes are called "statistically significant ($p<.05$)."[2]

The typical within-school standard deviation of the Composite score (the square root of 17.6) and the typical number of ACT-tested students at a high school (about 40) give us a hint about what to expect. Using tables of the Student *t* distribution, we can calculate that the critical value for the magnitude in a year-to-year difference in average ACT Composite score at a school with 40 students is about 1.9. This number is considerably larger than the typical magnitude of yearly change in average ACT Composite score (0.6 in this study). Therefore, we should expect that most observed yearly changes in high schools' average ACT Composite score are ambiguous.

---

[2] I avoid the use of hypothesis testing terminology. The null hypothesis of no systematic change is likely false in most instances, although the change might be very small.

It is important to note that an unambiguous change suggests only that there has been a change in the average level of achievement of different student cohorts. It does not prove that there has been a change in the effectiveness of instruction at the school. The reason is that change in achievement can result from many causes, including variables over which a school has no control. Of course, given an unambiguous observed change, it would be prudent to inquire whether it could have been influenced by changes in variables over which the school does have control.

Adjusting Observed Changes in Average Scores and Proficiency Rates

A change in the average score or proficiency rate at a high school can be reported as observed, with an associated confidence interval. These statistics address the question, "Has there been an unambiguous change (i.e., a change that is not plausibly due to chance) in the academic achievement of successive cohorts of students at the school?"

An observed change can also be statistically adjusted for concomitant changes in variables that are known to be related to academic achievement. Some of these covariates relate to student characteristics over which a high school typically has no control: Examples are background variables (gender, family income, race/ethnicity) and prior achievement in middle school. The statistically adjusted changes address the question, "Has there been an unambiguous change in the academic achievement of successive cohorts of students after taking into account changes in other student characteristics over which the school has no control?" In this study, I have calculated changes adjusted for certain background variables reported by students when they registered to take the ACT (see page 14). Some of the analyses also include prior achievement in middle school, as measured by the Explore Composite score (ACT, 2013). Prior achievement is usually the strongest predictor of current achievement (Sawyer, 2008; Sawyer &

Gibson, 2012). Of course, there are other variables related to ACT scores that are beyond the control of high schools, but that were not available for this study.

Note that an unambiguous change, even if adjusted for changes in variables over which the school has no control, still does not prove that there has been a change in the effectiveness of instruction at the school. The reason is that it is not usually feasible to collect data on all the important covariates; but, omitting these covariates in the models potentially biases the estimated changes.[3] Furthermore, adjusting for covariates places greater demands on student sample size which, as we shall see, is a strong limiting factor in accurately measuring change. Despite the limitations of cross-sectional changes, though, it would still be prudent after observing unambiguous adjusted changes to inquire whether they could have been influenced by changes in variables over which the school does have control.

In principle, one could also study the relationship between observed changes and variables over which a school has some influence, but not total control. Examples of such variables include students' attendance, behavior, and prior achievement in the same school (Sawyer, 2010). Proceeding further, one could study variables over which a school has considerable control (e.g., instructors and curriculum). I did not use variables like this in the study because data were not available for them.

[3] If it were possible, randomly assigning students to schools would mitigate some of the unobserved variable bias. Of course, students are rarely, if ever, randomly chosen by schools.

Research Questions

This study was intended to answer the following questions:

1. For what percentage of high schools can we detect an unambiguous change in ACT Composite score or attainment of all four ACT College Readiness Benchmarks (CRBs), using standard statistical procedures?
2. For what percentage of high schools can we detect an unambiguous change in ACT Composite score or all-CRB attainment, adjusted for background variables and prior achievement in middle school, using standard procedures?
3. Are there simple rules that high schools can apply to their observed changes that predict whether the changes are unambiguous?

The intent of the first question is to understand the big picture: Are the changes observed at most high schools large enough and based on large enough samples to be unambiguous, or can they instead plausibly be attributed to random variation in the students tested? The second question attempts a more nuanced understanding of the big picture, by examining whether adjusting observed changes for cohort differences in background characteristics and prior achievement changes the answer to the first question.

Most high schools do not have the resources to do formal statistical analyses to answer the first two questions. The intent of the third question is that schools might be able to do simple calculations that predict, with reasonable accuracy, whether a formal statistical analysis would find unambiguous changes. If a formal statistical analysis does confirm that the observed changes are unambiguous, the school could investigate potential reasons for the change.

Data

The outcome variables in this study are based on ACT Composite scores and on attainment of all four ACT CRBs. To answer the different research questions, I analyzed four separate analysis data sets that encompass different time spans and contain different sets of covariates:

- ACT-tested students (10-year sample): ACT student records from a random sample of 2,928 high schools with data from all of the graduating class years 2002 through 2011 (N=1,960,327 students).
- ACT-tested students (5-year sample): ACT student records from the same 2,928 schools, but only from the graduating class years 2007 through 2011 (N=1,080,843 students). The students represented in this file are a subset of the students represented in the previous file.
- Explore/ACT-tested students (10-year sample): Matched Explore/ACT records of students from all high schools with data from all of the graduating class years 2002 through 2011 (N=1,238 schools; 678,885 students).
- Explore/ACT-tested students (5-year sample): Matched Explore/ACT records of students from all high schools with data from all of the graduating class years 2007 through 2011 (N=2,613 schools; 703,786 students). The 1,238 schools in the Explore/ACT (10-year) file are a subset of the 2,613 schools represented in the Explore/ACT (5-year) file.

The ACT Composite scores in these data sets were principally obtained in grades eleven (38%) or twelve (62%), whenever students last took the ACT before graduating from high school. The Explore Composite scores were obtained in grade eight.

The 5-year and 10-year versions of the analysis data sets permitted studying the effect of time span on accurately measuring change. Because they are based on more data, changes over ten years should more frequently be unambiguous than changes over five years. ACT's summary reports to high schools currently include only five years of data, but by consulting its report from five years earlier, a high school could assemble information spanning ten years.

The ACT versions of the analysis data sets permitted studying the effect of using only background variables to adjust measures of change. The Explore/ACT versions of the analysis data sets permitted studying the effect of using both background variables and prior achievement (Explore Composite score) to adjust measures of change.[4] The adjusted measures of change might be useful when a high school experienced large changes in its students' background characteristics or prior achievement, as well as changes in its average ACT Composite scores. In this situation, the adjusted measures of change would provide support to the hypothesis that the changes in average ACT Composite scores were driven by changes in background characteristics or prior achievement.

Table 1 (pp. 12-13) summarizes the characteristics of the schools represented in the four analysis data sets, and Table 2 (pp. 14-15) summarizes the characteristics of the students. In both tables, these files are identified by the column headings "ACT-tested students (5-year sample)," "ACT-tested students (10-year sample)," "Explore/ACT-tested students (5-year sample)," and "Explore/ACT-tested students (10-year sample)." For simplicity, I refer to these files in text as ACT-5yr, ACT-10yr, Explore/ACT-5yr, and Explore/ACT-10yr, respectively.

In Table 1, the average number of students tested varies widely (from 1 to 940 in the ACT-5yr file, for example). Of course, a school with only one ACT-tested student per year

[4] An alternative analysis would study the difference between the ACT Composite score and the Explore Composite score as a measure of growth from grade eight to grades eleven/twelve.

should not attempt to estimate changes over time. Nevertheless, I retained all schools in the analysis in order to study the relationship between sample size and statistical precision over the full range of possible sample sizes.

In some states, all or substantially all public school students took the ACT test during the time period studied here:

- 2002 - 2011 graduating classes: Colorado and Illinois
- 2008 - 2011 graduating classes: Michigan
- 2009 - 2011 graduating classes: Kentucky, Wyoming
- 2010 - 2011 graduating classes: Tennessee
- 2011 graduating class: North Dakota

In other states, students decided themselves whether to take the ACT test. One would expect that, other things being equal, if the percentage of ACT-tested students in a school increases over time, then its average score would decline. A relevant high school characteristic in studying trends, therefore, is the percentage of ACT-tested students. The variable “Percent ACT-tested” in Table 1 was calculated by dividing the number of ACT-tested students in a particular graduating class by an estimate of the total twelfth-grade enrollment provided by Market Data Retrieval, Inc. Because the estimated twelfth-grade enrollment is constant over the time span indicated, rather than specific to particular graduating class years, the “Percent ACT-tested” in Table 1 is an approximation.

Table 1

Summary of School-Level Variables

School-level variable and quantile	Analysis data set				
	ACT-tested students (5-year sample)	**ACT-tested students (10-year sample)**	**Explore/ACT tested students (5-year sample)**	**Explore/ACT tested students (10-year sample)**	**ACT-tested students (2011 grad. class)**
Number of schools	2,928	2,928	2,613	1,238	24,860
Average number of students tested					
Minimum	1	2	1	1	1
Median	41	37	27	32	28
Maximum	940	883	680	538	973
Percent ACT-tested					
Minimum	1	1	<1	1	<1
Median	49	45	60	65	44
Maximum	100	100	96	91	100
Average Explore Composite score					
Minimum	*n.a*	*n.a*	9.0	8.0	*n.a*
Median	*n.a*	*n.a*	16.3	16.5	*n.a*
Maximum	*n.a*	*n.a*	25.0	24.0	*n.a*
Average ACT Composite score					
Minimum	12.9	13.0	9.0	12.0	6.0
Median	21.0	20.9	20.7	20.9	20.8
Maximum	31.4	31.5	35.0	34.0	35.0

(continued on next page)

Notes:

1. Quantiles of averages and percentages for the analysis data sets are calculated from data pooled over the indicated time span.
2. *n.a.* = not available.

Table 1 (continued)

Summary of School-Level Variables

School-level variable and quantile	**Analysis data set**				
	ACT-tested students (5-year sample)	**ACT-tested students (10-year sample)**	**Explore/ACT tested students (5-year sample)**	**Explore/ACT tested students (10-year sample)**	**ACT-tested students (2011 grad. class)**
Percent attaining all ACT CRBs					
Minimum	<1	<1	13	15	<1
Median	19	19	21	21	17
Maximum	93	94	31	31	100

Notes:
1. Quantiles of averages and percentages for the analysis data sets are calculated from data pooled over the indicated time span.
2. *n.a.* = not available.

Table 2

Summary of Student-Level Variables

	Analysis data set				
Student-level variable	**ACT-tested students (5-year sample)**	**ACT-tested students (10-year sample)**	**Explore/ACT tested students (5-year sample)**	**Explore/ACT tested students (10-year sample)**	**ACT-tested students (2011 grad. class)**
Number of students	1,080,843	1,960,327	703,786	678,885	1,623,112
Percent attaining all ACT CRBs	24	22	24	25	25
ACT Composite score					
Mean (SD)	21.1 (5.1)	21.1 (5.0)	21.4 (4.9)	21.8 (4.9)	21.1 (5.2)
Disability					
(percentage)	4	4	*n.a.*	*n.a.*	3
Explore Composite score					
Mean (SD)	*n.a.*	*n.a.*	16.5 (3.1)	16.7 (3.0)	*n.a.*
Family income					
(category percentages)					
< $36K	31	31	26	23	35
$36K - $60K	27	29	21	22	23
$60K - $80K	15	15	19	21	13
$80K - $100K	12	11	13	13	11
> $100K	16	14	21	21	18

(continued on next page)

Notes:

1. Percentages, means, and standard deviations are calculated from data pooled over the indicated time span.
2. *n.a.* = not available.

Table 2 (continued)

Summary of Student-Level Variables

	Analysis data set				
Student-level variable	**ACT-tested students (5-year sample)**	**ACT-tested students (10-year sample)**	**Explore/ACT tested students (5-year sample)**	**Explore/ACT tested students (10-year sample)**	**ACT-tested students (2011 grad. class)**
Gender					
(percent male)	45	45	46	46	46
Grade level at time of ACT testing					
(percent)					
11	38	38	40	35	40
12	61	62	60	64	59
Other	1	1	0	1	1
Parents' education index					
Mean (SD)	*n.a.*	*n.a.*	2.7 (1.0)	*n.a.*	*n.a.*
Race/ethnicity					
(category percentages)					
African-American	13	13	13	9	14
Asian / Pacific Islander	4	4	3	3	4
Hispanic or Latino	10	8	7	4	13
White	67	69	71	77	62
Other categories	7	7	7	7	7
U.S. citizen					
(percent)	97	97	98	99	96

Notes:
1. Percentages, means, and standard deviations are calculated from data pooled over the indicated time span.
2. *n.a.* = not available.
3. Parents' education index is defined as the sum of four dummy variables: mother completed high school, mother completed college, father completed high school, father completed college.

Table 2 shows the background variables considered in the analysis. These variables include: presence of a self-reported disability (either mental or physical), English primary language at home, family income, gender, grade level at time of ACT testing, parents' education index[5], race/ethnicity, and U.S. citizenship. Students self-reported values of these variables when they either took Explore or registered for or took the ACT.

Comparisons to the 2011 ACT Graduating Class File

Tables 1 and 2 both contain a column heading "ACT-tested students (2011 grad. class)." The data in this column enable us to study the representativeness of the four analysis data sets with respect to all students who took the ACT in 2011.

In Table 1, the median percentage ACT-tested, the median ACT Composite score, and the median percentage attaining all four ACT CRBs among schools represented in the ACT-5yr and ACT-10yr files are close to the corresponding medians in the 2011 graduating class file. The median number of students in the two ACT-tested files, however, is larger than the corresponding median in the 2011 graduating class file; the reason is that only larger schools consistently had data from all ten years 2002 - 2011.

In the two Explore/ACT-tested files, the median percentage ACT-tested is larger than the corresponding median in the 2011 graduating class file. The median average ACT Composite score is similar to the corresponding median in the 2011 graduating class file, but the median CRB attainment rate is somewhat higher.

Table 2 shows that the students represented in the Explore/ACT-5yr and Explore/ACT-10yr files are more likely than ACT-tested students in general to have high family incomes, to be

[5] Parents' education index was available only for the Explore/ACT-5yr file.

of white race/ethnicity, and to be U.S. citizens. The students represented in the two ACT-tested files are in most respects very similar to the ACT-tested students in the 2011 graduating class.

Multiple Imputation

Because the background variables are self-reported by students, they have missing values to varying extents. As shown in Table 3, the variables with the largest percentage of missing cases were family income for ACT-tested students (ACT-10yr file) - 24%, and parents' education index (Explore/ACT-5yr file) - 36%.[6]

Table 3

Percentage of Missing Values, Before Imputation, Among Student Background Variables

	Analysis data set		
Student background variable	**ACT-tested students (10-year sample)**	**Explore/ACT tested students (5-year sample)**	**Explore/ACT tested students (10-year sample)**
Disability	8	0	*n.a.*
Family income	24	14	3
Gender	1	0	0
Grade level at time of ACT testing	1	0	0
Parents' education index	*n.a.*	36	*n.a.*
Race/ethnicity	6	0	0
U.S. citizen	4	3	3

Note: *n.a.* = not available.

[6] Table 3 does not contain a column for ACT-tested students (5-year sample), because these students are a subset of the ACT-tested students (10-year sample).

Simply excluding cases with any missing values can potentially introduce bias in the results. To reduce the potential for bias, I imputed missing values using SAS PROC MI. To properly estimate confidence intervals with imputed data, one should repeat the analyses on multiple imputations of the original data set. In previous studies, I have found that confidence interval widths typically increase only slightly after including among-imputation variance. To simplify the analyses, therefore, I based results on only one imputation.

Method

The analyses in this study estimated changes over time in the average ACT Composite score and all-CRB attainment rate of individual high schools. The changes pertain to time spans that education officials and journalists can study, given the summary reports ACT produces each year.

For each comparison over time, I calculated a "change variable" representing the comparison:

- *Diff2yr*, a dummy variable equal to 1 for students in the last graduating class (2011 in these data) and equal to 0 for students in the previous year's graduating class (2010 in these data)
- *Trend5yr*, a linear sequence variable reflecting a linear trend over the last five years (2007-2011)

Diff2yr represents a change from year to year; *Trend5yr* represents a trend over the preceding five years. I also calculated change variables based on ten years of graduating classes:

- *DiffL5mF5*, a dummy variable distinguishing the last five years' graduating classes from the first five years' graduating classes.
- *Trend10yr*, a linear sequence variable reflecting a linear trend over the last ten years.

The last two change variables can help us learn the extent to which comparisons based on ten years are more frequently unambiguous than when based on five years.

Each change variable served as a predictor in a regression model for predicting the ACT Composite score (or probability of all-CRB attainment) of individual students at a particular school. For basic comparisons (not taking into account background characteristics or prior achievement), the change variable was the only predictor in the regression model. For making adjusted comparisons, I also included variables representing background characteristics and prior achievement.

In the ACT Composite score models, the weight associated with a change variable reflected a school's change in average score for the time period indicated.[7] To illustrate, if the weight for the change variable *Diff2yr* is 0.1 at a particular high school, then its average Composite score for the 2011 graduating class is estimated to be 0.1 units larger than its average Composite score for the 2010 graduating class. A weight of 0.1 for *Trend5yr* suggests that the average Composite score increased by 0.1 unit per year from the 2007 graduating class to the 2011 graduating class.

School officials can make analogous approximate comparisons from the high school summary reports produced by ACT. Replicating the analyses in this study would require student-level records, however. High schools and school districts can purchase electronic files of their students' ACT records for a nominal fee.

Model Development

Average ACT Composite score. Changes for the average ACT Composite score were estimated from linear regression models with various covariates. For the basic comparisons (not

---

[7] The weights for the models estimating all-CRB attainment are also related to changes in a school's all-CRB attainment rate, but through an extra calculation described later.

adjusted for background variables or prior achievement), the only covariate was a change variable. For changes adjusted for potential background or prior achievement covariates, I included the covariates listed in Table 3, in addition to the dummy or linear sequence change variable, in the linear model. To select the covariates, I estimated parsimonious models (i.e., those for which each covariate's weight was statistically significant ($p<.05$) at half or more of the high schools).[8] The models with background or prior achievement covariates always included the time dummy variable or linear sequence change variable, regardless of whether its estimated weight suggested an unambiguous change (i.e., its 95% confidence interval did not include the value 0). I did not include any interaction effects in the models.

I also estimated models with the school-level variable *Percent ACT-tested*. This variable was not statistically significant ($p<.05$) at half or more of the high schools for any of the models. I did not include any other time-varying school-level variables in the models.

All-CRB attainment. I estimated changes related to all-CRB attainment rate in a similar way, using logistic models instead of linear models. In the logistic model, the outcome variable is dichotomous (0=did not attain all CRBs; 1=attained all CRBs). The predictor variables were the time dummy variable or linear sequence change variable, as well as any covariates. From the 95% confidence interval for the weight of the change variable, I determined whether the associated change was unambiguous.

The weights in the logistic model are not equal to the estimated changes as they are in the linear model, but they can be used to calculate the changes. The logistic model yields for each student an estimate of the student's log-odds ($\widehat{lo}$) of attaining all CRBs. The estimated log-odds

[8] In principle, one could construct a parsimonious model separately for each high school, but that would be difficult and time-consuming. To simplify building the models, I stratified the high schools in each analysis data set on the average number of ACT-tested students, and then estimated parsimonious models separately by stratum.

$\widehat{lo}$ is a linear combination of the change variable and values of the covariates. It can be converted to an estimated probability for each student:

$$\hat{p} = 1/[1 + \exp(-\widehat{lo})] \tag{1}$$

I calculated the average $\hat{p}$ over all students in a school, given the relevant value of the change variable and the values of any covariates, as an estimate of the all-CRB attainment rate for a given year.

At some small high schools, either no students attained all of the CRBs, or all students attained all of the CRBs, during the relevant time period. It is not possible using standard methods to fit logistic models to data like these. Therefore, I removed these schools' data from the analysis. As a result, the number of schools decreased as follows:

- ACT-5yr analysis, from 2,928 to 2,114
- ACT-10yr analysis, from 2,928 to 1,886
- Explore/ACT-5yr analysis, from 2,613 to 1,879
- Explore/ACT-10yr analysis, from 1,238 to 758.

The high schools removed from the analysis were smaller than the other schools. Most had low average ACT Composite scores, although a few had high average ACT Composite scores.

Hierarchical models. I also investigated hierarchical versions of all of the linear models, with random effects associated with the change variables. Hierarchical models like these reflect the structure of the data, in which students are nested within graduating class years. Hierarchical models could, therefore, in principle yield more accurate estimates of the change variable weights. I found, however, that the estimated variances associated with the change variables were not statistically significantly different from zero ($p<.05$) at a majority of high schools for any of the linear models. Expecting similar results for the logistic models, I did not attempt to estimate hierarchical versions of them. Therefore, all of the results in this report pertain to standard fixed-effects models.

One could also consider estimating hierarchical models that reflect the nesting of students within schools, or the cross nesting of students within schools and graduating class years. Because individual high schools do not have access to the data needed to estimate such models (namely, the data from the other schools), I did not include these other types of hierarchical models in this study. The feasibility of developing school district or state summary reports based on hierarchical models with crossed effects is an interesting idea to consider for future research.

Summary Statistics

I summarized the distribution over schools of the weights corresponding to the change variables. These statistics tell us how often we can expect to observe changes of different directions and magnitudes.

Each change variable and each high school had an associated regression model. For each change, I calculated the percentage of schools for which the corresponding weight was unambiguous (i.e., the 95% confidence interval did not include the value 0).[9] These percentages provide the information to answer research questions 1 and 2.

Flag Variables

Although the statistical procedures used in this study are standard, school officials typically do not have the resources to perform them. Another goal of this study, therefore, was to create flag variables that officials could easily calculate from their summary reports and that predict whether the corresponding difference or trend is unambiguous. A value of the flag

[9] Confidence intervals for changes associated with average ACT Composite score were calculated from the Student *t* distribution. Confidence intervals for changes associated with all-CRB attainment were calculated from the chi-square distribution.

variable above a certain cutoff would suggest a reasonable chance that the associated change is unambiguous. School officials could then decide whether to do a formal statistical analysis to confirm the result.

Ideally, school officials would keep track, over a period of years, of changes in their students' entering characteristics, the local environment, and their school improvement efforts. If the flag variables suggested an unambiguous change, and a formal statistical analysis confirmed the result, officials could undertake a more thorough investigation to determine whether any of the previously documented changes in students' entering characteristics, local environment, and school improvement efforts were related to the changes in average test scores or CRB attainment rates. Changes in students' entering characteristics, local environment, and school improvement efforts could also be documented retrospectively, although this would probably be more difficult and less accurate.

Whether a two-year difference is unambiguous depends on the magnitude of the difference, on the size of the student sample from which it was calculated, and on the variation of the scores. The flag variable for the two-year difference is based on the first two of these quantities, the magnitude of the difference and student sample size. Let $\overline{\mathrm{X}}_i$ be the average ACT Composite score, and let N_i be the number of students tested, from year *i*. The *t* statistic used to calculate the 95% confidence interval for *Diff2yr* = $\overline{\mathrm{X}}_{2011} - \overline{\mathrm{X}}_{2010}$ is proportional to $\sqrt{\frac{N_{2010}N_{2011}}{N_{2010}+N_{2011}}(N_{2010}+N_{2011}-2)}\ (\overline{\mathrm{X}}_{2011} - \overline{\mathrm{X}}_{2010})$. To simplify users' calculations, I defined the flag variable for *Diff2yr* as $\frac{N_{2010}+N_{2011}}{2}\ (\overline{\mathrm{X}}_{2011} - \overline{\mathrm{X}}_{2010})$.

- For the difference variables *Diff2yr* and *DiffL5mF5*, the corresponding flag variables are the average sample size over the relevant years multiplied by the observed difference in average scores.

- For *Trend5yr*, the flag variable is the average five-year sample size multiplied by $(\overline{X}_{2011} - \overline{X}_{2007})/4$.
- For *Trend10yr*, the flag variable is the average ten-year sample size multiplied by $(\overline{X}_{2011} - \overline{X}_{2002})/9$.

I calculated flag variables for change and each high school.

To determine whether the flag variables usefully predict their associated changes, I estimated from the results of all high schools a logistic regression model with the following outcome variable:

Y = 1, if the school's observed difference or trend is unambiguous

= 0, otherwise.

The predictor variable in each model was the corresponding flag variable. From the estimated model, I calculated the cutoff value of the flag variable for which y = 1 at 50% or more of the high schools (if such a value exists). I then calculated the following percentages:

1. The percentage of all schools whose value of the flag variable is above the cutoff
2. Among schools below the cutoff, the percentage for which $y = 1$
3. Among schools above the cutoff, the percentage for which $y = 1$
4. Among schools for which y = 1, the percentage above the cutoff.

Percentage 1 tells us the percentage of schools whose flag variable suggests an unambiguous change. Percentages 2 and 3 show the extent to which the percentage of schools with unambiguous changes depends on whether their flag variables are above or below the cutoff. A useful flag variable should result in a low value of Percentage 2 and a high value of Percentage 3. Percentage 4 shows, among the schools with unambiguous change, how many are captured by the flag variable and cutoff.

I also defined analogous flag variables for the changes related to all-CRB attainment, using all-CRB attainment rates instead of average ACT Composite scores. The mathematical correspondence between the all-CRB attainment flag variables and the *t* statistics for the associated changes is less direct than for the ACT Composite score analyses. I investigated whether the flag variables for all-CRB attainment would still be usefully predictive.

Creating flag variables adjusted for students' background characteristics or prior achievement would negate the purpose of creating flag variables (namely, simplicity). Therefore, the flag variables are defined only for unadjusted changes.

Results

In the preceding section, the terms *Diff2yr*, *Trend5yr*, *DiffL5mF5*, and *Trend10yr* refer to the change variables corresponding to different comparisons over time. In this section, these terms also refer to the weights and flag variables corresponding to the different comparisons.

Covariates for Adjusting Estimated Changes

Table A-1 in the appendix shows the covariates in the parsimonious models for adjusting *Trend5yr* and *Trend10yr* related to average ACT Composite score. The covariates are listed separately by analysis data set and by the average number of ACT-tested students between 2002 and 2011.

Table A-1 shows that there were more background variables in models based on the ACT-5yr and ACT-10yr samples than in models based on the Explore/ACT-5yr and Explore/ACT-10yr samples. The reason is that in models based on Explore/ACT-tested students, prior achievement in middle school (as measured by Explore Composite score) was a very strong predictor.

In models for adjusting the two-year difference *Diff2yr*, there were few student-level covariates. In models estimated from the ACT-5yr and ACT-10yr samples, family income was the only covariate. In models estimated from the Explore/ACT-5yr and Explore/ACT-10yr samples, Explore Composite score was the only covariate.

Table A-2 shows the covariates in the parsimonious models for adjusting *Trend5yr* and *Trend10yr*, as calculated from all-CRB attainment. This table is organized similarly to Table A-1. There are fewer covariates listed in Table A-2 than in Table A-1, reflecting the greater demand that logistic models place on sample size. As was noted earlier, models based on the two analysis data sets for ACT-tested students had more covariates than models based on the two analysis data sets for Explore/ACT-tested students, because prior achievement was a very strong predictor.

Distribution of Estimated Changes

Table A-3 in the appendix summarizes the distributions, over high schools, of the unadjusted changes related to average ACT Composite score. The medians show that at typical high schools, there was little change over time, no matter how change was measured. This result is consistent with that reported by Ziomek (2000). On comparing the top and bottom sections of Table A-3, we see that the medians of changes based on data from ACT-tested students were very similar to the medians for the two Explore/ACT-tested files.

The minima, first quartiles, third quartiles, and maxima in Table A-3 show that increasing the time span in a comparison reduced its variability. Figure 1 on the following page illustrates this result by comparing the distributions of the magnitudes (absolute values) of *Diff2yr*, *Trend5yr*, and *Trend10yr*. Note that many more schools had large magnitudes of *Diff2yr* than of *Trend5yr*, and many more schools had large magnitudes of *Trend5yr* than of *Trend10yr*. In other words, changes based on short time spans more frequently took on large values than did changes based on longer time spans. This result is expected, because year-to-year fluctuations average out over longer time periods.

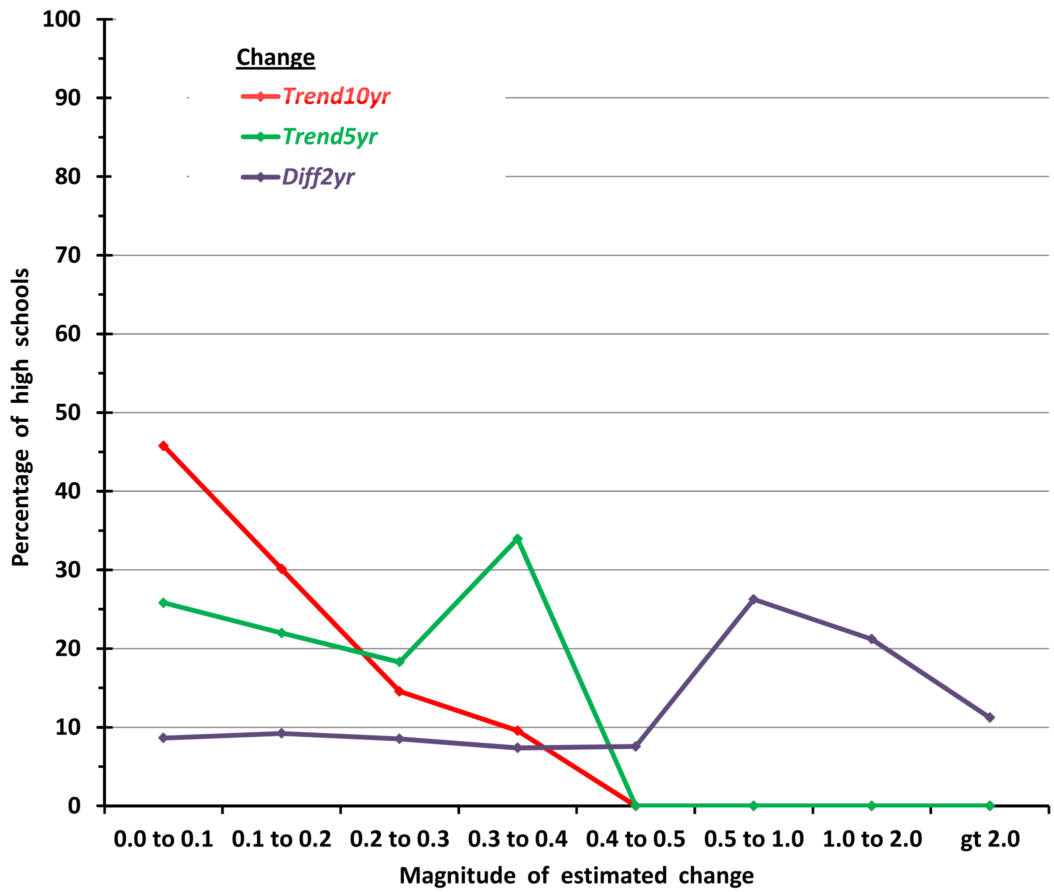


Figure 1. Percentage of high schools, by magnitude of *Diff2yr*, *Trend5yr*, and *Trend10yr* related to average ACT Composite score.

Table A-4 in the Appendix contains analogous information about the marginal distributions of the adjusted changes related to average ACT Composite score. For the ACT-tested students, the distributions in Table A-4 were very similar to the corresponding distributions of unadjusted changes in Table A-3. The medians for the Explore/ACT-tested students in Table A-4, however, were slightly larger than the corresponding medians for the

Explore/ACT-tested students in Table A-3. The differences in the medians for the Explore/ACT-tested students show the effects of adjusting for prior achievement.

As in Table A-3, the minima, first quartiles, third quartiles, and maxima in Table A-4 show that increasing the time span in a comparison reduced its variability. For example, the maximum values of *Diff2yr*, *Trend5yr*, and *Trend10yr* for the Explore/ACT-tested students in Table A-4 are 6.9, 2.9, and 0.8, respectively.

Tables A-5 and A-6 in the Appendix summarize the marginal distributions of the unadjusted and adjusted changes related to all-CRB attainment rate. The changes related to all-CRB attainment rates showed the same general pattern as the changes related to average ACT Composite score. At typical high schools, there was little change over time, no matter how change was measured. Many more schools had large magnitudes of *Diff2yr* than of *Trend5yr*, and many more schools had large magnitudes of *Trend5yr* than of *Trend10yr*. Finally, the distributions of changes based on data from ACT-tested students were very similar to the distributions based on the two Explore/ACT-tested files.

The following sections discuss results related to the percentage of high schools whose changes were unambiguous (i.e., whether they could plausibly be attributed to systematic changes over time in the achievement of different student cohorts). I first discuss unambiguous changes related to average ACT Composite score, then present a parallel discussion related to all-CRB attainment.

Unambiguous Changes Related to Average ACT Composite Score

Table 4 (see following page) shows the percentage of high schools with unambiguous changes related to average ACT Composite score. The table is organized by data source (ACT-tested students or Explore/ACT-tested students), by the type of change, and by whether the change was adjusted for covariates.

Table 4 is based on data from all high schools in each of the four analysis data sets, including very small high schools and high schools with small values of the changes. The section beginning on p. 31 shows how the percentage of high schools with unambiguous results is related to the number of ACT-tested students and to the magnitude of the changes. The section beginning on p. 34 shows how the percentage of high schools with unambiguous results is related to "flag" variables (defined as the product of the number of ACT-tested students and the magnitude of the changes).

As one would expect, the percentage of high schools with unambiguous changes increased sharply with the number of years on which the changes were based. About one-third of high schools had unambiguous values of *Trend10yr*. Less than one-tenth of schools had unambiguous values of the two-year differences.

Table 4

Percentage of Schools with Unambiguous Changes over Time in Average ACT Composite Score, by Data Source, Change, and Adjustment for Student Covariates

	ACT data		**Explore/ACT data**	
Change	**No student covariates**	**With student covariates**	**No student covariates**	**With student covariates**
Diff2yr	9	8	12	12
DiffL5mF5	33	29	23	37
Trend5yr	21	17	16	11
Trend10yr	36	32	28	34

The relationship to data source and to adjustment by covariates was more complex. For the unadjusted changes, a greater percentage of the schools in the ACT-tested files than in the Explore/ACT-tested files had unambiguous changes, if based on five or more years of data. For example, 36% of the unadjusted *Trend10yr* values from the ACT-10yr file were unambiguous, compared to 28% of the unadjusted Trend10yr values in the Explore/ACT-10yr file. This difference might be due to the larger within-school sample sizes in the two ACT-tested files than in the two Explore/ACT-tested files (see Table 1).

A complication in interpreting the unadjusted changes based on the ACT data is that during the period 2002 - 2011, some school districts and states began census-testing their students. As the pool of students tested became larger, one would expect unadjusted average scores to decline. As a result, more of these high schools might have had unambiguous declines in their unadjusted average scores than if their districts or states had not census-tested. On the other hand, adopting census ACT testing is likely to have had less effect at the Explore/ACT schools, because a larger percentage of their students were ACT-tested to begin with.

For the adjusted changes, high schools in the Explore/ACT-tested files tended to have unambiguous changes somewhat more frequently than schools in the ACT-tested files. For example, 34% of the adjusted *Trend10yr* values from the Explore/ACT-10yr file were unambiguous, compared to 32% of the schools from the ACT-10yr file. One plausible explanation is that although the distribution of adjusted changes in the EXPLORE/ACT-tested schools was similar to the distribution in the ACT-tested schools, the prior achievement covariate in the Explore/ACT-tested group (Explore Composite score) removed much of the variation among student cohorts that was unaccounted for in the models based on the ACT-tested files.

Stated another way, in the ACT-5yr and ACT-10yr samples, adjusting changes with student background variables tended to decrease the frequency of unambiguous results. In the Explore/ACT-5yr and Explore/ACT-10yr samples, however, adjusting changes with both student background variables and prior achievement tended to increase the frequency of unambiguous results.

Relationship to number of students tested and magnitude of change. Whether a change is unambiguous depends on, among other things, both the sample size on which it is based and its magnitude. Figures 2 and 3 show these relationships for changes, unadjusted for student background characteristics, as calculated from the ACT-10yr and ACT-5yr samples.

Figure 2 on the following page shows the percentage of schools with unambiguous changes, by number of students tested, for the various changes unadjusted for student background characteristics. In Figure 2, the size of each plot symbol corresponds to the number of high schools in each category of the horizontal axis (average number of ACT-tested students). For example, there were many more schools with 50 or fewer ACT-tested students than with 250 or more ACT-tested students.

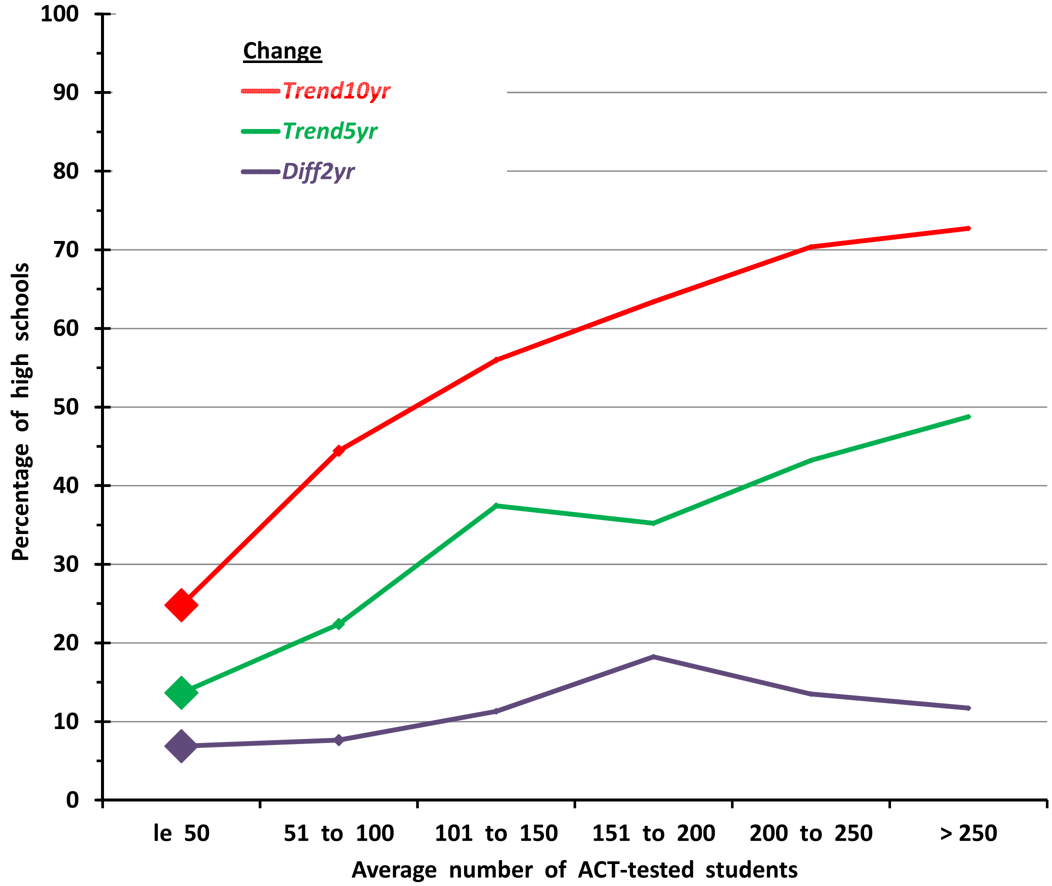


Figure 2. Percentage of high schools with unambiguous two-year difference, five-year trend, or ten-year trend related to average ACT Composite score, by average number of ACT-tested students.

In Figure 2, the percentage of schools with unambiguous *Trend5yr* or *Trend10yr* increases with sample size, but only the *Trend10yr* line ever exceeds 50 percent. The line for *Diff2yr* never exceeds 20 percent. Moreover, the *Trend10yr* line is strictly increasing with sample size, but the *Trend5yr* and *Diff2yr* lines are not. The most likely reason for this result is that trends estimated from five or fewer years of data are more susceptible to random fluctuations than trends based on ten years of data. Another possibility, of course, is that the expected values of *Trend5yr* and *Diff2yr* are not strictly increasing with sample size.

Figure 3 shows comparable percentages by the magnitude of the various changes. As in Figure 2, the size of each plot symbol corresponds to the number of high schools in each category of the horizontal axis.

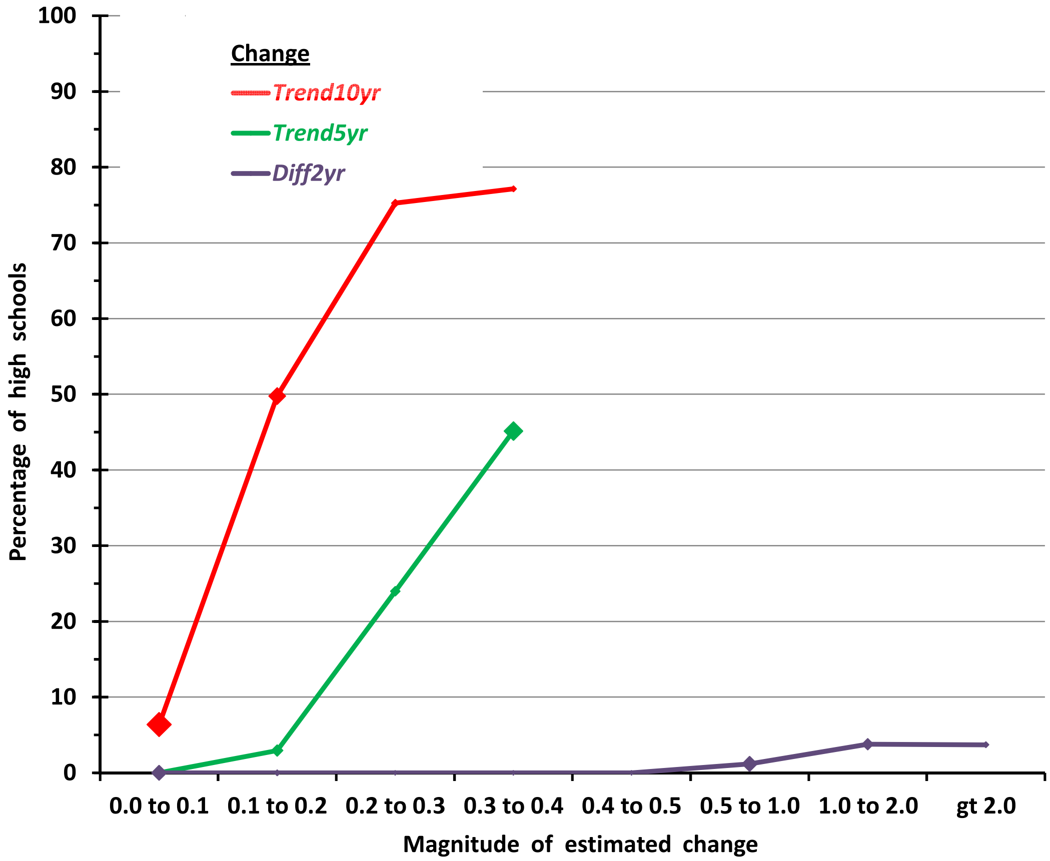


Figure 3. Percentage of high schools with unambiguous two-year difference, five-year trend, or ten-year trend related to average ACT Composite score, by magnitude of change.

The percentage of high schools with unambiguous *Trend10yr* increased sharply as the observed values of these trends increased. In contrast, even very large values of *Diff2yr* were rarely unambiguous. There were no schools with *Trend5yr* or *Trend10yr* magnitudes above 0.4.

Figures 2 and 3 are based on changes calculated from the ACT-5yr and ACT-10yr analysis data sets, and are unadjusted for student background characteristics. The changes adjusted for student background characteristics, although not shown in these figures, had similar relationships with sample size and change magnitude. The changes calculated from the Explore/ACT-5yr and Explore/ACT-10yr files also had similar relationships with sample size and change magnitude.

Flag variables. Figure 4 illustrates the relationship between values of the flag variables (horizontal axis) and the percentage of high schools whose corresponding changes were unambiguous (vertical axis). The lines are based on data in the ACT-5yr and ACT-10yr files.

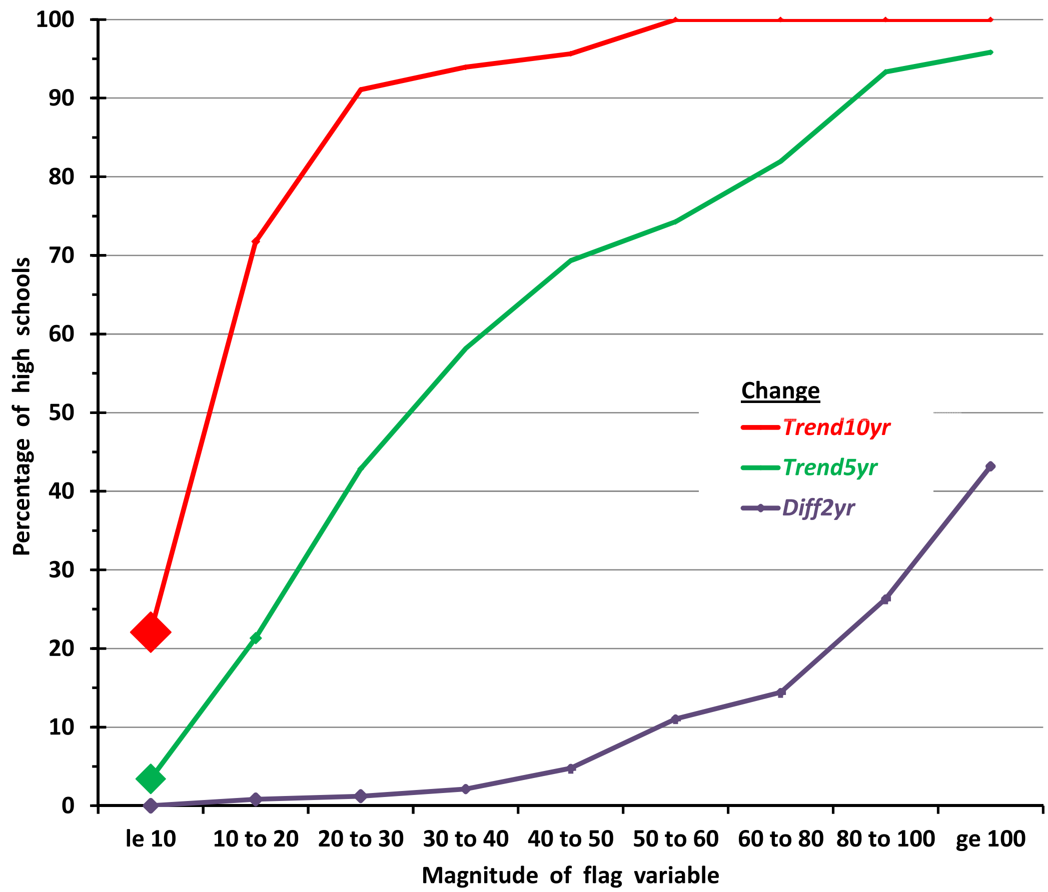


Figure 4. Percentage of high schools with unambiguous two-year difference, five-year trend, or ten-year trend related to average ACT Composite score, by magnitude of flag variable.

Figure 4 shows that the flag variable for *Trend10yr* was very predictive of whether the estimated *Trend10yr* change was unambiguous. The problem, though, is that for 76% of high schools, the *Trend10yr* flag variable was less than or equal to 10, as indicated by the large red plot symbol. The *Trend5yr* flag variable also predicted whether that change was unambiguous, although its line is less steep than that for *Trend10yr*. Moreover, the modal category for the Trend5yr flag variable was also less than or equal to 10. The line for the flag variable for *Diff2yr* never reaches 50%, indicating that it is not as useful as the other two flag variables.

Table 5 on the next page shows the cutoffs and accuracy statistics for all the flag variables. For example, the flag variable for *Diff2yr* has a cutoff of 166; this means that the average ACT Composite score of a high school with 100 ACT-tested students would have to change by 1.66 score units to be flagged. Only 4 percent of all schools exceeded the cutoff of 166, and the flag variable captured only about 32% of all unambiguous two-year differences. In contrast, the cutoff for the *DiffL5mF5* flag is 43. At a school with 100 students, the corresponding value of *DiffL5mF5* is .43 score units, and about 27% of schools exceeded the cutoff. Moreover, the *DiffL5mF5* flag captured about 70% of all the unambiguous values of *DiffL5mF5*. Thus, predictions for *Diff2yr* are less useful than those for *DiffL5mF5*.

Table 5

Flag Variable Analysis for Predicting Unambiguous Changes Related to Average ACT Composite Score

			Percent unambiguous		
Change	**Cutoff for flag variable**	**Percent above cutoff**	**Below cutoff**	**Above cutoff**	**Capture percent**
Diff2yr	166	4	6	66	32
DiffL5mF5	43	27	14	86	70
Trend5yr	34	13	13	76	47
Trend10yr	10	24	22	81	54

The same pattern pertains to the flag variables for trends. A greater percentage of schools exceeded the cutoff for the *Trend10yr* flag than for the *Trend5yr* flag. The *Trend5yr* flag captured slightly less than half of all unambiguous values of *Trend5yr*, and the *Trend10yr* flag captured slightly more than half of all unambiguous values of *Trend10yr*.

Unambiguous Changes Related to All-CRB Attainment

Table 6 (see following page) shows the percentage of high schools with unambiguous changes related to all-CRB attainment. The table is organized by data source (ACT-tested students or Explore/ACT-tested students), by the type of change, and by whether the change was adjusted for covariates.

The results in Table 6 for all-CRB attainment parallel those for average ACT Composite score (Table 4), but show lower percentages of unambiguous changes, either positive or negative. The lower percentages in Table 6 resulted from the greater demands on sample size by the logistic regression models (Equation (1)). About 31% of high schools had unambiguous unadjusted values of *Trend10yr*. Only about 6% of schools had unambiguous unadjusted values of *Diff2yr*. One would expect that at approximately 5% of schools the confidence interval for *Diff2yr* would not include the value 0, even if the "true" value of *Diff2yr* were 0 at all schools. Given the number of schools in the analysis (2,114), the 6% result here is not inconsistent with an assumed value of 0 for *Diff2yr*.

Table 6

Percentage of Schools with Unambiguous Changes over Time in All-CRB Attainment, by Data Source, Change, and Adjustment for Student Covariates

	ACT data		**Explore/ACT data**	
Change	**No student covariates**	**With student covariates**	**No student covariates**	**With student covariates**
Diff2yr	6	*n.a.*	7	*n.a.*
DiffL5mF5	28	25	18	27
Trend5yr	15	13	10	14
Trend10yr	31	27	23	33

As in Table 4, the unadjusted changes from the two ACT-tested files were more frequently unambiguous than were the unadjusted changes from the two Explore/ACT-tested files. The adjusted changes from the Explore/ACT-tested files were more frequently unambiguous than were the adjusted changes from the ACT-tested files

Relationship to number of students tested and change size. Figure 5 shows the percentage of schools with unambiguous changes in all-CRB attainment, by number of students tested, for the three changes *Diff2yr*, *Trend5yr*, and *Trend10yr*. The statistics are based on data from the ACT-5yr and ACT-10yr files, and are unadjusted for student background characteristics. The size of the plot symbols is roughly proportional to the number of high schools in the associated sample size categories.

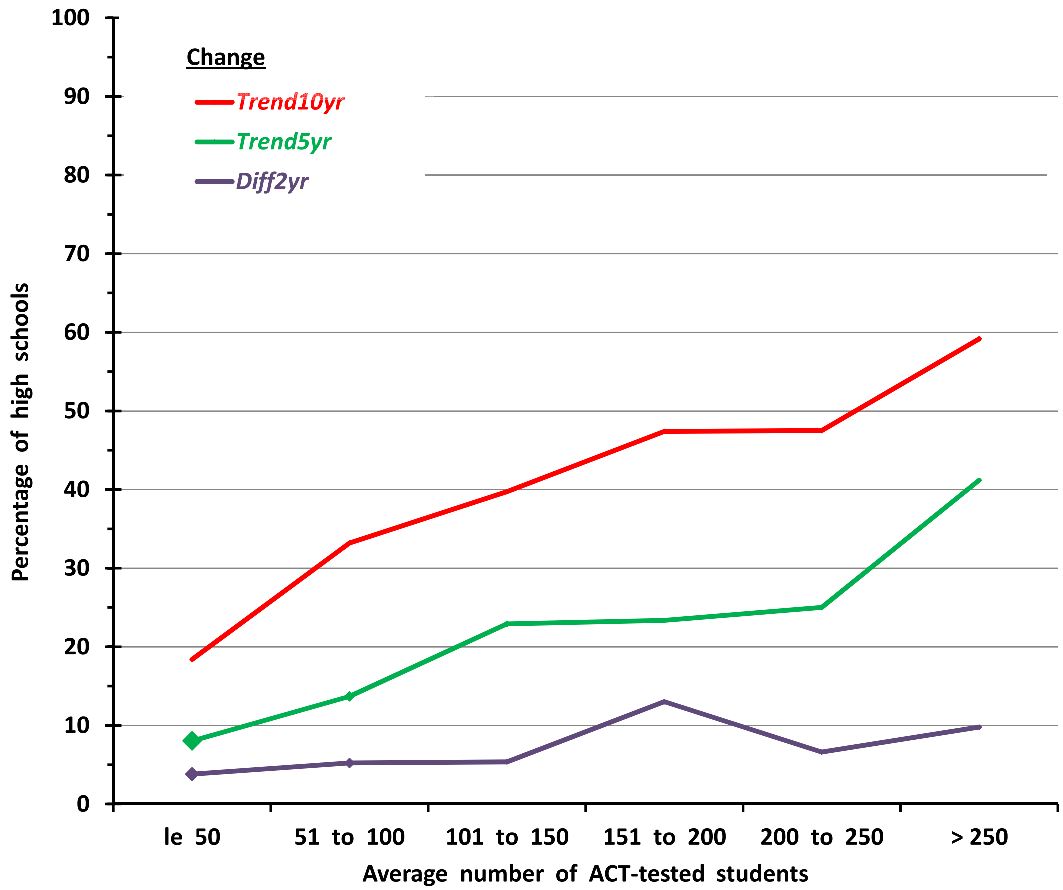


Figure 5. Percentage of high schools with unambiguous two-year difference, five-year trend, or ten-year trend related to all-CRB attainment, by average number of ACT-tested students.

As one would expect, the percentage of schools with unambiguous *Trend5yr* or *Trend10yr* increases with sample size, but only the *Trend10yr* line ever exceeds 50 percent. The line for *Diff2yr* never exceeds 20 percent.

Figure 6 shows comparable percentages by the magnitude (absolute value) of the various changes.

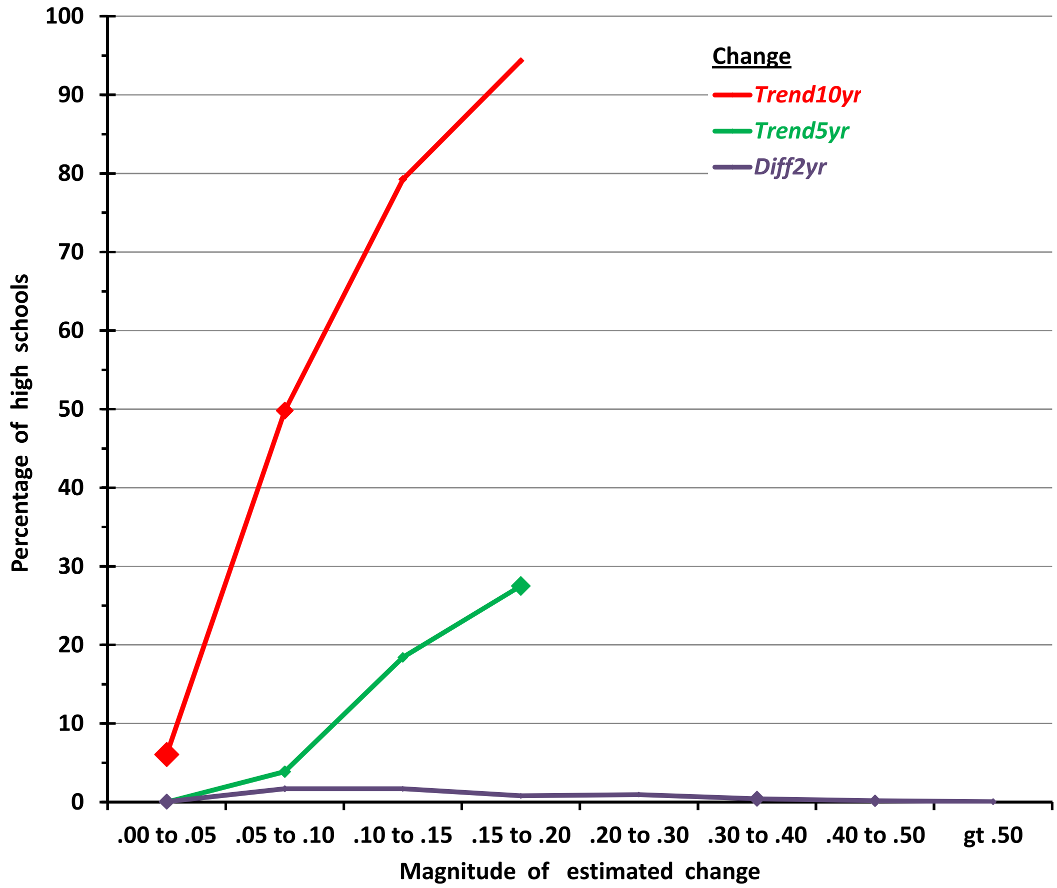


Figure 6. Percentage of high schools with unambiguous two-year difference, five-year trend or ten-year trend related to all-CRB attainment, by magnitude of change.

Note that the percentage of high schools with unambiguous *Trend10yr* and *Trend5yr* increases as the observed values of these trends increase. In contrast, even very large values of *Diff2yr* are rarely unambiguous.

Although not shown in Figures 5 and 6, the changes adjusted for student background characteristics had similar relationships with sample size and change magnitude. The changes calculated from the Explore/ACT-5yr and Explore/ACT-10yr files also had similar relationships with sample size and change magnitude.

Flag variables. Figure 7 illustrates the relationship between flag variables (horizontal axis) and the percentage of schools whose corresponding changes were unambiguous (vertical axis). The statistics are based on data in the ACT-5yr and ACT-10yr files.

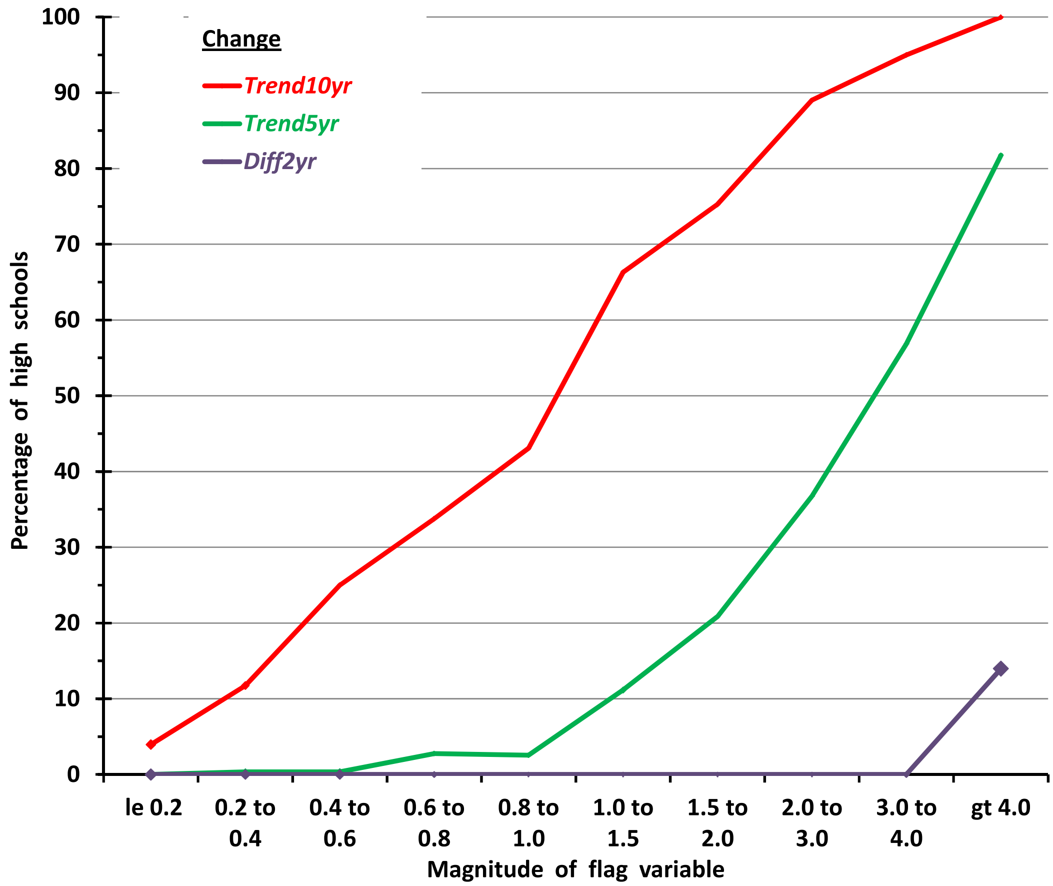


Figure 7. Percentage of high schools with unambiguous two-year difference, five-year trend, or ten-year trend related to all-CRB attainment, by magnitude of flag variable.

Figure 7 shows that the flag variable for *Trend10yr* predicted well whether *Trend10yr* was unambiguous. The problem is that for most high schools, the flag for this variable was small. The flag variable for *Trend5yr* also predicted whether its associated change was unambiguous, although its line was less steep than that for *Trend10yr*. The line for the flag variable for *Diff2yr* never reaches 20%, indicating that it was not useful.

Table 7 below shows the cutoffs and accuracy of the flag variables for predicting unambiguous changes in all-CRB attainment. Predictions for changes based on five years of data were less useful than those based on ten years of data. The predictions for *Diff2yr* were not useful: Only 3% of schools exceeded the cutoff, and the flag captured only 30% of the schools with unambiguous values of *Diff2yr*.

Table 7

Flag Variable Analysis for Predicting Unambiguous Changes Related to All-CRB Attainment

			Percent unambiguous		
Change	**Cutoff for flag variable**	**Percent above cutoff**	**Below cutoff**	**Above cutoff**	**Capture percent**
Diff2yr	17.7	3	4	65	30
DiffL5mF5	5.5	22	12	83	67
Trend5yr	3.3	9	8	75	47
Trend 10yr	1.1	22	17	79	56

Discussion

The principal goal of this study was to determine whether changes over time in the average ACT Composite scores or College Readiness Benchmark (CRB) attainment rates at most high schools unambiguously suggest systematic differences among the different cohorts of students, or whether instead they are plausibly due to random variation. Another goal of the study was to develop flag variables that could easily be calculated from data in high school summary reports and that predict whether formal statistical analyses would find that the changes over time are unambiguous.

At most schools, changes over time in average ACT Composite scores and all-CRB attainment rates are plausibly attributed to chance, at least according to the standard statistical procedures applied in this study. This result pertains to all of the changes, whether based on two, five, or ten years of data. The ten-year trend in average ACT Composite score is the most frequently unambiguous (at 36% of high schools). The five-year trend in average ACT Composite score is unambiguous at 21% of high schools. In contrast, the two-year difference in average ACT Composite score is unambiguous at only 9% of high schools. Adjusting average ACT Composite scores for student background characteristics or prior achievement changes the percentages somewhat, but does not change the general conclusion: Changes over time in average scores and attainment rates are plausibly attributed to chance at most high schools.

The fundamental reason is that changes in average ACT Composite score or all-CRB attainment rate are too small at most schools, given their sample size, to be unambiguous. For example, at the typical high school in the sample, there were approximately 45 ACT-tested students. The flag variable analysis suggests that at schools of this size, an average change of about .22 per year in average ACT Composite score, sustained over 10 years, is needed for a ten-year trend to be unambiguous. A two-year difference of about 3.69 in average ACT Composite score is required for it to be unambiguous. Most schools did not have changes of this magnitude.

Another way to state this result is that time is not an important predictor of ACT Composite score or all-CRB attainment. In contrast, ACT Composite score and all-CRB attainment are associated with background characteristics and prior achievement at most high schools. This result is already well-known, but it suggests that at most high schools, characteristics that have formed in the past are more important than changes that have occurred recently in driving students' academic achievement. Another potential implication is that a long time frame is typically needed to accurately detect changes in student readiness. Sustained long-term approaches to school improvement may be more likely than “quick fixes” to result in detectable improvement in student achievement (ACT, 2012b).

At some high schools, of course, changes are unambiguous and should be investigated further. It is worth reiterating, though, that an unambiguous trend or difference does not by itself prove that there has been a change in the instructional effectiveness of a school. There are many variables other than instructional effectiveness that influence students’ achievement as measured by test scores. An unambiguous change in test scores should therefore not be interpreted as a conclusion about instructional effectiveness, but instead as a suggestion to search for explanations. Some of the explanations could involve variables beyond the control of schools, while others, such as curriculum and instruction, are within the control of schools.

Many schools do not have the resources to do the statistical analyses in this study. Working from the summary reports currently produced by ACT, however, high school officials could calculate flag variables that accurately predict, at least for some of the changes, whether they are unambiguous. If the flag variable as calculated for a particular high school exceeds the cutoff in Tables 5 or 7, then officials could engage statisticians to do formal analyses. The results in this study suggest that the flag variable approach would be useful for studying five-year trends and ten-year trends, but not for studying two-year differences.

Another reasonable response to the findings in this study would be for testing companies to provide in their summary reports to high schools confidence intervals for changes, as well as guidance on how to interpret the confidence intervals. Users could then identify unambiguous differences and trends. Doing this would reduce over-interpretation of changes that are plausibly explained by random variation among different student cohorts.

References

ACT. (2007). The ACT technical manual. Iowa City, Iowa: Author. Retrieved October 15, 2013 from http://www.act.org/aap/pdf/ACT_Technical_Manual.pdf.

ACT. (2012a). ACT national profile report. Iowa City, Iowa: Author. Retrieved May 16, 2013 from http://www.act.org/newsroom/data/2012/pdf/profile/National2012.pdf.

ACT. (2012b). The core practice framework: A guide to sustained school improvement. Iowa City, Iowa: Author. Retrieved August 30, 2013 from http://www.act.org/research/policymakers/pdf/Core-Practice-Framework-Guide.pdf .

ACT. (2013). Explore technical manual. Iowa City, Iowa: Author. Retrieved June 10, 2013 from http://www.act.org/explore/pdf/TechManual.pdf.

Allen, J. (2013). Updating ACT's College Readiness Benchmarks (ACT Research Report Series 2013 (6)). Iowa City, Iowa: ACT. Retrieved June 26, 2013 from http://www.act.org/research/researchers/reports/pdf/ACT_RR2013-x.pdf.

Allen, J., & Sconing, J. (2005). Using ACT Assessment Scores to Set Benchmarks for College Readiness (ACT Research Report Series 2005-3). Iowa City, Iowa: ACT. Retrieved June 10, 2013 from http://www.act.org/research/researchers/reports/pdf/ACT_RR2005-3.pdf.

College Board. (2012). SAT national profile report. New York: Author. Retrieved May 16, 2013 from http://media.collegeboard.com/homeOrg/content/pdf/sat-report-college-career-readiness-2012.pdf.

Davidson, E., Reback, R., Rockoff, J., & Schwartz, H. (2013, April). Fifty ways to leave a child behind: Idiosyncrasies and discrepancies in states' implementation of *NCLB*. Paper presented at the Association for Education Finance and Policy conference. Retrieved May 17, 2013 from http://www.columbia.edu/~ekd2110/Fifty_Ways_4_5_2013.pdf.

McNeil, M. (2013). Waiver states under scrutiny. *Education Week*, *33*(2), 1. Retrieved October 16, 2013 from http://www.edweek.org/ew/toc/2013/08/28/index.html.

Nebraska Department of Education. (2012). Nebraska Performance Accountability System. Lincoln, NE: Author. Retrieved May 17, 2013 from http://www.education.ne.gov/assessment/pdfs/Nebraska_Performance_Accountability_System_Aug_2012.pdf.

Sawyer, R. (2008). Benefits of Additional High School Course Work and Improved Course Performance in Preparing Students for College (ACT Research Report Series 2008-1). Retrieved June 6, 2013 from http://www.act.org/research/researchers/reports/pdf/ACT_RR2008-1.pdf.

Sawyer, R., & Gibson, N. (2012**).** Exploratory Analyses of the Long-Term Effects of Improving Behavior, Attendance, and Educational Achievement in Grades 1–6 and 8–12 (ACT Research Report Series 2012 (3). Retrieved June 6, 2013 from http://www.act.org/research/researchers/reports/pdf/ACT_RR2012-3.pdf.

Twing, J. (2013). Growth, gains, and progress under NCLB. Pearson Educational Measurement. Retrieved May 16, 2013 from www.pearsonschool.com/index.cfm?locator=PSZ1Az.

Wyatt, J., Kobrin, J., Wiley, A., Camara, W., & Proestler, N. (2011). SAT Benchmarks (Research Report 2011-5). New York: College Board. Retrieved June 10, 2013 from http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2011-5-sat-college-readiness-benchmark-secondary-performance.pdf.

Ziomek, R. (2000). Monitoring changes in high school average ACT Composite scores over time (ACT Information Brief 2000-2). Iowa City, Iowa: ACT. Retrieved July 8, 2013 from http://www.act.org/research/researchers/briefs/2000-2.html.

Appendix

Table A-1

Student Covariates in the DiffL5mF5, Trend 5yr, and Trend10yr ACT Composite Score Models

	Analysis data set			
Average number of students tested	**ACT-tested students (5-year sample)**	**ACT-tested students (10-year sample)**	**Explore/ACT-tested students (5-year sample)**	**Explore/ACT-tested students (10-year sample)**
1-75	Family income	Family income	Explore Composite Family income	Explore Composite Family income
76-300	African-American Disability Family income Grade level Hispanic	African-American Disability Family income Grade level Hispanic	African-American Explore Composite Parents' education index	African-American Explore Composite Family income
≥ 301	African-American Asian Disability Family income Grade level Hispanic	African-American Asian Disability Family income Grade level Hispanic	African-American Explore Composite Family income Parents' education index U.S. citizen	African-American Explore Composite Family income U.S. citizen

Table A-2

Student Covariates in the DiffL5mF5, Trend5yr, and Trend10yr All-CRB Attainment Models

	Analysis data set			
Number of students tested	**ACT-tested students (5-year sample)**	**ACT-tested students (10-year sample)**	**Explore/ACT-tested students (5-year sample)**	**Explore/ACT-tested students (10-year sample)**
1-75		Family income	Explore Composite	Explore Composite
76-300	Family income	African-American Family income Gender	Explore Composite	Explore Composite
≥ 301	African-American Disability Family income Gender Grade level Hispanic	African-American Disability Family income Gender Grade level Hispanic	Explore Composite Family income	Explore Composite Family income

Table A-3

Distribution, over High Schools, of Unadjusted Changes Related to Average ACT Composite Score, by Data Source

Change	Min.	Q1	Med.	Q3	Max.
ACT-tested students					
Diff2yr	-8.8	-0.6	0.0	0.7	13.2
DiffL5mF5	-4.8	-1.0	0.2	0.8	5.1
Trend5yr	-2.4	-0.2	0.0	0.2	2.1
Trend10yr	-0.8	-0.1	0.0	0.1	0.9
Explore/ACT-tested students (5-year sample)					
Trend5yr	-3.8	-0.3	0.0	0.2	2.8
Explore/ACT-tested students (10-year sample)					
Diff2yr	-8.9	-1.9	0.0	0.8	9.0
DiffL5mF5	-4.2	-0.6	0.1	0.6	4.6
Trend10yr	-0.8	-0.1	0.0	0.1	0.7

Table A-4

Distribution, over High Schools, of Adjusted Changes Related to Average ACT Composite Score, by Data Source

Change	**Min.**	**Q1**	**Med.**	**Q3**	**Max.**
ACT-tested students					
Diff2yr	-10.7	-1.5	0.0	0.6	25.0
DiffL5mF5	-5.1	-0.3	0.1	0.7	4.7
Trend5yr	-2.5	-0.2	0.0	0.2	1.7
Trend10yr	-0.9	-0.1	0.0	0.1	0.8
Explore/ACT-tested students (5-year sample)					
Trend5yr	-2.4	-0.1	0.1	0.3	2.9
Explore/ACT-tested students (10-year sample)					
Diff2yr	-14.0	-1.1	0.1	0.6	6.9
DiffL5mF5	-2.7	-0.1	0.3	0.7	3.8
Trend10yr	-0.7	0.0	0.1	0.1	0.8

Table A-5

Distribution, over High Schools, of Unadjusted Changes Related to All-CRB Attainment Rate, by Data Source

Change	Min.	Q1	Med.	Q3	Max.
ACT-tested students					
Diff2yr	-.39	-.05	.00	.05	.51
DiffL5mF5	-.29	-.01	.02	.06	.32
Trend5yr	-.11	-.01	.01	.02	.14
Trend10yr	-.04	.00	.00	.01	.06
Explore/ACT-tested students (5-year sample)					
Trend5yr	-.17	-.01	.00	.02	.14
Explore/ACT-tested students (10-year sample)					
Diff2yr	-.47	-.05	.00	.05	.51
DiffL5mF5	-.24	-.01	.02	.05	.49
Trend10yr	-.05	.00	.00	.01	.07

Table A-6

Distribution, over High Schools, of Adjusted Changes Related to All-CRB Attainment Rate, by Data Source

Change	Min.	Q1	Med.	Q3	Max.
ACT-tested students					
Diff2yr	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*
DiffL5mF5	-.29	-.01	.02	.06	.33
Trend5yr	-.11	-.01	.01	.02	. 14
Trend10yr	-.05	.00	.00	.01	.07
Explore/ACT-tested students (5-year sample)					
Trend5yr	-.24	-.01	.00	.02	.17
Explore/ACT-tested students (10-year sample)					
Diff2yr	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*
DiffL5mF5	-.25	-.01	.02	.06	.51
Trend10yr	-.08	.00	.00	.01	.07