

High-Stakes Academic Assessment & Educational Evaluation in Illinois

Dr. Steve Cordogan
Director of Research and Evaluation
Township High School District 214
steve.cordogan@d214.org
website: www.d214.org/research

High-stakes standardized tests are vital for evaluating students and schools.

- Currently, they are the best way to compare academic between schools and districts.
- They provide a reality test to balance the typical school's perception that they are above average (the "Lake Wobegon" syndrome).
- They can provide invaluable information for school improvement.

But reducing the assessment of school and student academic quality to a few test scores also can be highly inaccurate.

- Prior classroom performance is almost always a better predictor of a student's future classroom performance than test scores.
- Future college and career performance is the best measure of school/district quality.
- Tests are a better measure of all students in a school than of individual students.
- Demographics largely determine academic performance
- Demographics largely determine growth on vertically aligned standardized tests

More Qualifiers

- Data fluctuate randomly between years.
- Magnitude of school improvement across years is dependent on initial quality - initially underperforming schools will look better.
- Magnitude of school improvement across years is dependent on demographic consistency.
- Some versions of the same test are harder than others.
- Some tests are better than others.

All testing organizations and publishers of performance comparisons have one thing in common: they are businesses seeking to stay in business and make money.

- Testing companies are sellers of test instruments, and should be trusted only as much as any other business.
- Announcements of annual test performance by testing organizations may have some academic purpose, but also are part of a marketing strategy.
 - Negative headlines are almost always more newsworthy than positive ones.
- The primary purpose of magazines and newspapers who publish school ratings is to sell magazines and newspapers.
- There is a lot of bad data out there. Some of the best-publicized data often are the worst.

Published Data Can Be Very “Misleading”

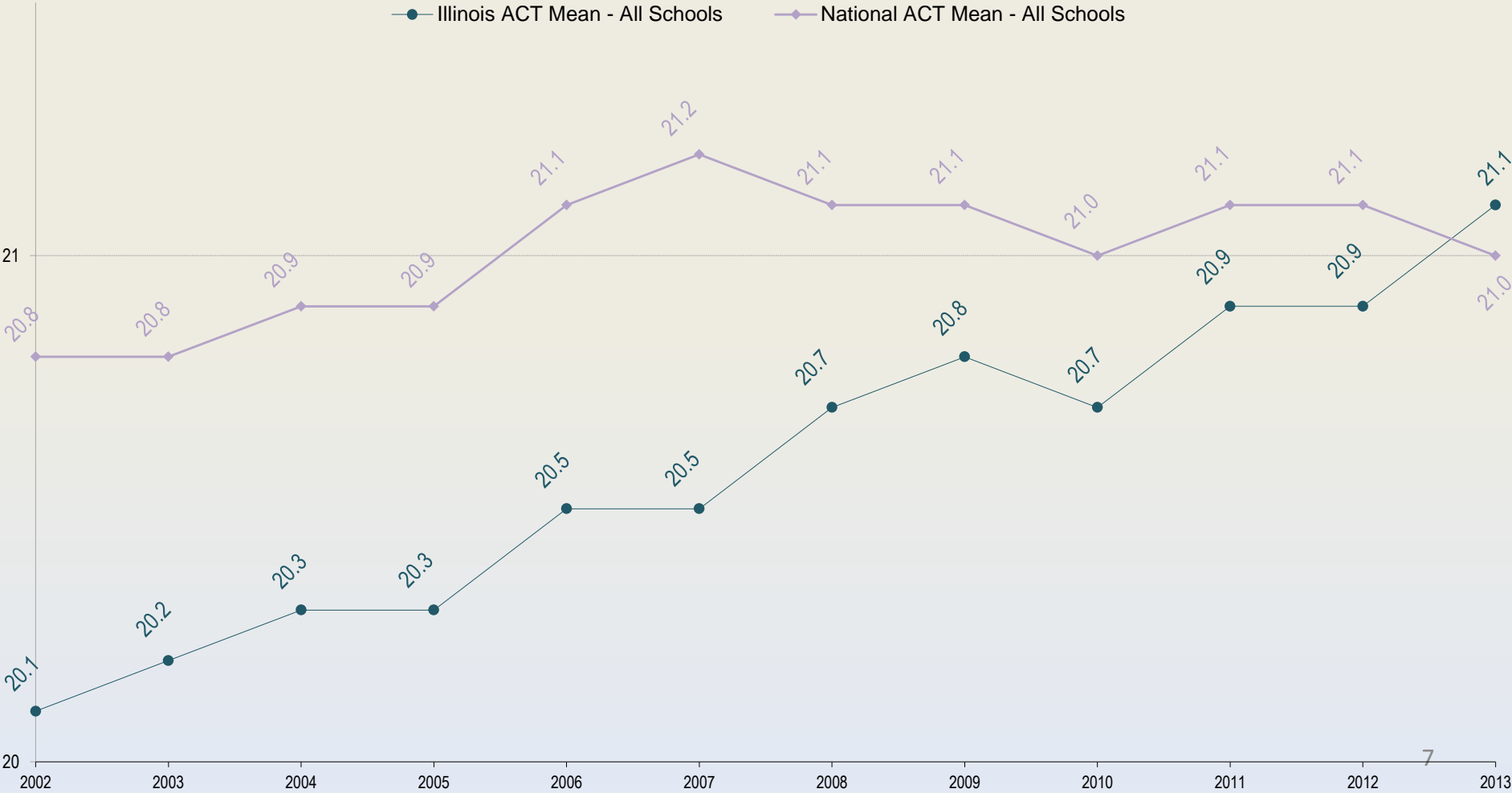
Headline: “Illinois ACT scores post biggest drop in a decade” Chicago Tribune, August 21, 2013

But in the sub-headline: “Testing company now including scores of specially accommodated students in results, which dragged Illinois down”

The reality: ACT scores in Illinois reached a 12 year record high (since universal testing began).

The Actual Comparison Shows Something Very Different from the Headline

ACT Composite Score Performance in Illinois by Graduating Class



ACT Has Been a Leader in Promoting the Illusion of an Educational Crisis

- ACT maintains that only 25% of the high school students in the U.S. are ready for college.
- The 25% figure is a product of over-interpreting its weakest ACT subject tests and indefensible methodology.
- ACT's own studies refute the figure.
- The state report card's use of the percentage of students scoring above the national composite score average is much more realistic.

Standards should form the basis for curriculum and assessments. Current Illinois academic standards are:

- 1997 Illinois State Standards (being replaced by Common Core Standards)
- ACT College Readiness Standards
- Common Core Standards - for Illinois, 44 other states, Washington D.C., and four territories

Major High Stakes Tests: Undergraduate College Placement

- ACT (originally American College Testing)
- Scholastic Aptitude Test (SAT)
- Scholastic Aptitude Test - subject tests
- ACT COMPASS (community college focus)
- ACCUPLACER (community college focus)
- Advanced Placement (AP) test

Major High Stakes Tests: Current State-Mandated High School Evaluation

- Prairie State Achievement Examination (PSAE) - for almost all students, including English Language Learners
 - ACT
 - WorkKeys workplace readiness assessment
 - State science test

The PSAE soon will be replaced by PARCC or ASPIRE.
- Illinois Alternate Assessment (IAA) - for highly cognitively disabled students
- ACCESS - language proficiency (not subject area content) for English Language Learners (ELL)
- Alternate ACCESS - for highly cognitively disabled students

Major High Stakes Tests: Non-Mandated

Current High School Evaluation

- Educational Planning and Assessment System (EPAS).
 - ACT college placement test
 - PLAN (a scaled-down ACT for late 9th grade and 10th grade admission)
 - EXPLORE (a scaled-down ACT for 8th and 9th grade admission)

The PLAN and EXPLORE will soon to be replaced by ASPIRE.

- Preliminary Scholastic Aptitude Test (PSAT) - also for National Merit Scholarship Program
- Advanced Placement (AP) test

Major High Stakes Tests: K-8 School Evaluation

- Illinois Standards Achievement Test (ISAT) - soon to be replaced by PARCC or ASPIRE
- EXPLORE (to be replaced by ASPIRE)
- Measures of Academic Progress (MAP)
- Iowa Tests of Basic Skills (ITBS)
- Cognitive Abilities Test (COGAT)

Major High Stakes Tests: Future School Evaluation

- ACT ASPIRE (a single test that will replace both PLAN and EXPLORE, and add tests for grades 3 - 8)
- Partnership for Assessment of Readiness for College and Careers (PARCC) – future grades 3-11 assessment system for Illinois, 14? other states, Washington D.C., and the Virgin Islands
- Smarter Balanced Assessment Consortium - future assessment system for 22 other states

Positive Aspects of Staying with ACT

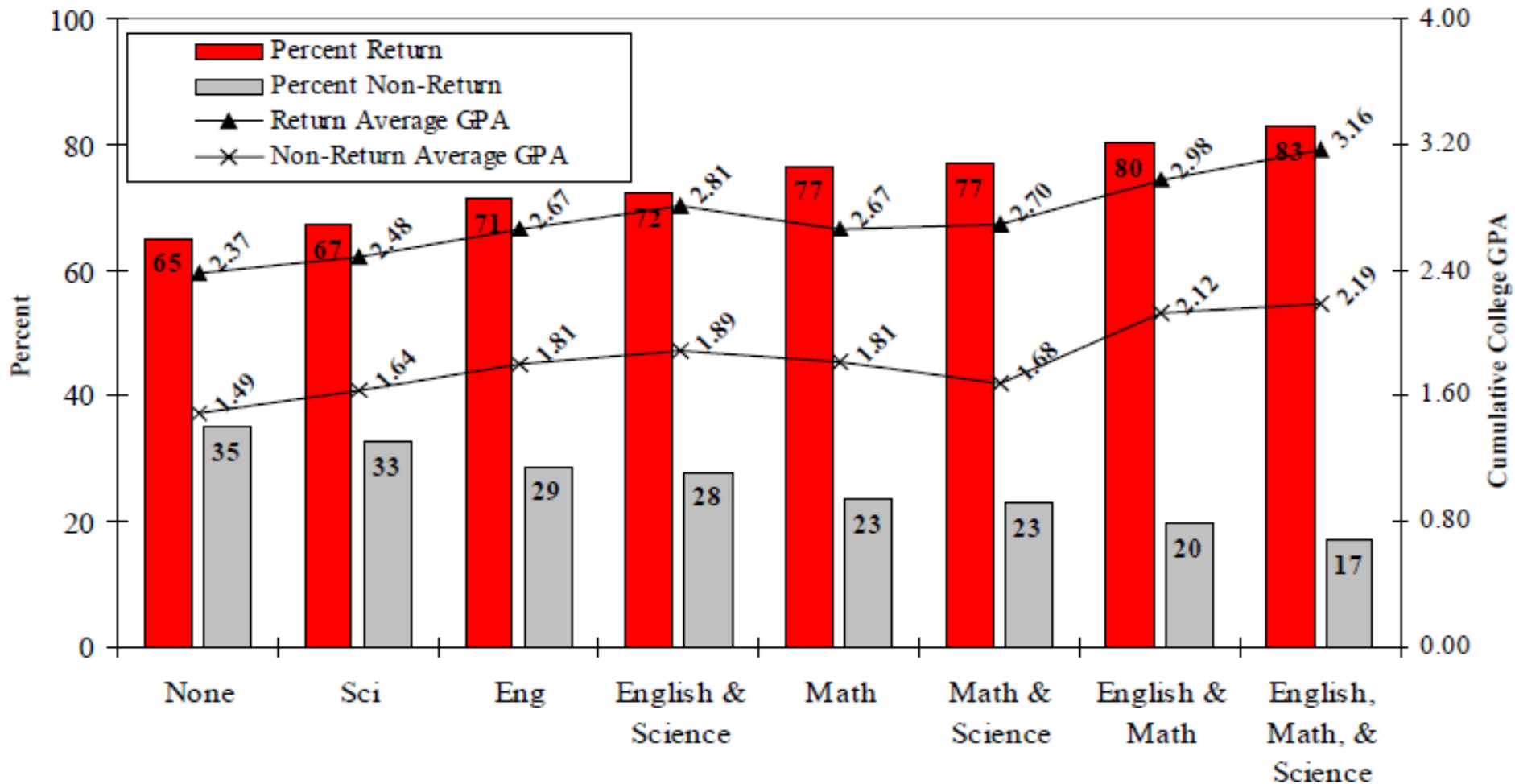
- ACT tests have been useful for school improvement and initial diagnosis of student ability.
- We already know the ACT test.
- The ACT test is meaningful to our students.

Negative Aspects of Staying with ACT

- ACT science and reading tests are of very limited value.
- The degree to which the ACT products align to the common core is limited.
- ASPIRE, due to its relative brevity, may not adequately measure high school student progress.
- Grade 3 through 8 ASPIRE tests are an unknown.
- ACT does not seem to be ready to deliver ASPIRE tests.
- ACT has required Illinois to administer the never-scientificallly-validated WorkKeys suite of tests.
- THE EXPLORE and PLAN replacement tests are much more expensive than in prior years.
- Newer products from ACT have been disappointing.
- ACT's corporate ethics have been questionable.

ACT Knows its Limitations, but Has Not Made the Changes Needed to Fix Them

Figure 1: Public Four-Year Retention and College GPAs by ACT College Benchmarks



Positive Aspects of Going with PARCC

- Tests focus on mathematics and language arts, the two most accurately testable areas.
- Longer tests mean greater reliability and validity
- Tests were designed from the Common Core and should best represent them.

Negative Aspects of Going with PARCC

- Missed deadlines, defecting states, and an apparent lack of readiness to deliver the test pilot does not inspire confidence.
- All of their tests are an unknown.
- The tests are so much longer than current ones that they make unrealistic demands on a school's schedule.
- Administration window timing is very problematic.
- Students currently have minimal motivation to perform well on the tests.
- Most high school students will continue to need to take the ACT until/if the PARCC test can be used for college admissions.

The Move to More Authentic Testing

- Both PARCC and ACT are trying to produce more authentic tests, moving away from multiple choice tests to short answer and problem solving.
- Such testing potentially is more authentic, since it can require higher-level thinking and a more complete picture of student understanding.
- But such tests are inherently less reliable because they involve subjectivity in the grading of the student response.
- Such tests only are of superior validity if the grader takes the time and has the skill to grade the questions consistently and accurately.

National and International Comparison Tests

- National Assessment of Educational Progress (NAEP) - national testing conducted by National Center for Educational Statistics (4th, 8th, and 12th grade mathematics, reading, science, writing, the arts, civics, economics, geography, U.S. history, and Technology and Engineering Literacy)
- Program for International Student Assessment (PISA) - international testing conducted by National Center for Educational Statistics (15-year-old students' reading, mathematics, and science literacy).
- Trends in International Mathematics and Science Study (TIMSS) - international testing affiliated with National Center for Educational Statistics (Grades 4 and 8)

The Limits to National and International Comparison Tests Findings

- The sampling procedures are not remotely random (they usually test and therefore do not provide comparable data between states and countries).
- The lack of comparability is greatest in international tests, where large percentages of the children in many countries, specifically the poorer and less educated ones, are excluded from testing (and often from schooling in general).
- Motivation is critical to student performance. Our over-tested students have little motivation to perform well on these tests. The problem is greatest at the higher grade levels.

Publishers of Annual or Occasional School Ratings:

- Washington Post
- Newsweek
- US News and World Report
- Chicago Magazine

Other Sources for School Rating Data:

- ACT Annual Final Graduating Class Score Release/College Readiness Benchmark Performance
- The ACT High School to College Success Report
- School Report Card Data
- Illinois Interactive Report Card

So What Are the Best Data Sources?

- Test Data
 - ACT scores
 - Number of passing AP tests per student or % of graduates passing at least one AP test.
 - Improvement on a test from year to year (e.g., ACT scores across several years of graduating classes)
 - Growth from EXPLORE to PSAT ACT
 - MAP data
- College Data
 - College attendance levels
 - College GPA
 - College graduation levels
 - Quality of colleges from which degrees are obtained

Where Can We Get College Data?

- The ACT High School to College Success Report
- High school exit surveys
- High school follow-up surveys
- Illinois P-20 database
- National Student Clearinghouse data

How Well Can We Evaluate Teachers with Test Data?

High stakes test data can provide insight into a teacher's performance but no quantitative system of teacher evaluation can provide sufficiently accurate data on teacher performance for systematic personnel decisions.

- The test instruments would have to be validated and carefully matched to the course.
- There would have to be enough different sections and teachers for meaningful contrasts.
- The number of students per class would need to be sufficiently large.
- The demographics of the students and initial performance levels would have to be used to adjust expectations.
- Performance in prior and concurrent related courses also would have to be considered (e.g., math impacts physics).²⁷

In Summary:

- The thoughtful use of assessment data has improved student learning.
- Student performance has been improving, but they have not been reported in the media, and the inclusion of more academically at-risk students sometimes has obscured the improvements
- Much good academic assessment data exist, but there also are bad data everywhere.
- Even good data must be used carefully, or they will provide bad information.
- Many current tests provide very useful data, though significantly better tests are within reach.
- The next few years are critical, since the test choices we make now will impact student learning for years to come.²⁸