



Why Scores on the ACT® Test Are Scores You Can Trust

WAYNE CAMARA, Senior Vice President, Research

DEBORAH HARRIS, Chief Research Scientist

Score scales are what we use to report test scores earned by individual students as well as groups of students. The score scale is also what allows students, educators, colleges, and scholarship agencies to interpret and make informed use of test scores. When a new test is introduced, it can take years for users to develop a deep knowledge and understanding of what the scores mean in terms of student achievement—what a student with a given score can and cannot do. And if the score scale is unstable, users may never be able to understand what scores mean.

The current ACT 1 to 36 score scales for English, reading, and science were established in 1988 and implemented in 1989 with the introduction of the revised ACT® test. Over 100,000 high school students were tested to provide a nationally representative sample to conduct the scaling. Allowing students to use calculators in 1996 had a psychometric impact on the scores, requiring math test scores to be rescaled at that time. Various stability and validity studies conducted by ACT over the years have shown the scales to be consistent over forms and over years in terms of both the meanings of the reported scores and in their ability to predict meaningful outcomes, such as college course grades.

Because of the consistency of ACT score scales, we have been able to develop the ACT College and Career Readiness Standards (CCRS), which are empirically derived descriptions of the essential skills and knowledge students must possess to become ready for college and career. The CCRS give clear meaning to test scores by providing the knowledge and skills students scoring in particular score ranges are likely to possess, and they serve as a link between what students have learned and what they are ready to learn next.

The stability of the score scales has also allowed ACT to identify the ACT College Readiness Benchmarks—scores on the ACT subject tests that represent the level of achievement required for students to have a 50% chance of obtaining a B or higher or about a 75% chance of obtaining a C or higher in corresponding credit-bearing first-year college courses.

These uses of score scales allow students to interpret what their scores mean in a meaningful context—what a 23 indicates about the knowledge and skills a student can demonstrate, and whether the student is likely to succeed in postsecondary coursework.

Changes are integral to the relevance of the ACT—tweaks primarily based on results from comprehensive curriculum surveys of educators from elementary schools through college. These changes are made incrementally and thoughtfully, allowing the interpretation of the scores scales to remain consistent and letting policymakers and educators compare the performance of groups of students (for example, freshman classes, graduating high school students, and states) across years. Comparing group performance trends is essential to monitoring changes in the college and career readiness of students over multiple years and is only possible when the constructs and content measured by a test remain fundamentally similar from year to year, as they are with the ACT. When significant changes are introduced to a test, such trends cannot be maintained, and the scores cannot be fairly compared across years.

The two best examples of such breaks in trends can be traced to the National Assessment of Educational Progress (NAEP) and the SAT. For example, NAEP notes: “Although long-term trend and main NAEP both assess mathematics and reading, there are several differences, particularly in the content assessed, how often the assessment is administered, and how the results are reported. These and other differences mean that results from long-term trend and main NAEP cannot be compared directly.”¹

Similarly, when the College Board introduced content changes and eliminated antonyms from the SAT Verbal Test in 1995, it had to recenter the SAT scale to allow users to link pre- and post-1995 SAT scores. For example, a pre-1995 SAT Verbal score of 730 corresponds to a post-1995 score of 800. This type of recentering was also an example of a break in the score scale; colleges, schools, states, policymakers, and the media that track group performance over time could not simply compare scores. Instead, they needed to try to convert the old scores to the new scale using complex and time-consuming statistical translations before they could compare the scores.

If the score scale is unstable, users may never be able to understand what scores mean.

¹ “What Are the Differences Between Long-Term Trend NAEP and Main NAEP?” National Assessment of Educational Progress (NAEP), last updated June 25, 2013, http://nces.ed.gov/nationsreportcard/about/ltr_main_diff.aspx.

As you can imagine, disruptions in the score scale can be especially difficult for college admissions officers who must evaluate applicants for admission based on test scores that were obtained across years where the scores are not equivalent or for educators who are attempting to track changes in aggregate scores over time. The SAT recentering created confusion about how to fairly compare students who completed the test in 1994 (on the old scale) and 1995 (on the new scale). Such a disruption or break in the scale is likely for the upcoming revised SAT given the dramatic changes announced in content, constructs, and item types.

It is important for those who use test scores to understand the appropriate uses of scores and be aware of disruptions or changes to score scales. It is the commitment to keeping our score scales consistent and stable—to avoiding dramatic revisions and changes—that helps make scores on the ACT scores you can trust. ■

Wayne J. Camara is Senior Vice President of Research at ACT. His research background spans the ACT continuum of solutions. Before he came to ACT, he held positions at the American Psychological Association and College Board, where he served as vice president for research and development.

Deborah J. Harris is Chief Research Scientist at ACT. She provides psychometric consultation and support for resident and contract programs and ensures the effective delivery and communication of measurement and reporting services. She is also an adjunct professor at the University of Iowa and has authored or coauthored over 140 publications and presentations.

ACT[®]