

**Accountability Considerations When Using a College-Readiness Assessment as Part of a  
School Accountability System:**

**A Technical and Policy Brief from the Advisory Committee to the Select Committee on  
Statewide Education Accountability<sup>1</sup>**

**Judy Catchpole, Chair**

**November 4, 2016**

The Advisory Committee to the Select Committee on Statewide Education Accountability had recommended to the Select Committee to use the state-administered college readiness assessment as part of the readiness indicator in the school accountability system (WAEA) in order to use the assessment for the purpose for which it was designed. Following this line of reasoning, the Advisory Committee further recommended to eliminate the use of the college readiness exam as an achievement indicator. This follows the 2015 recommendations of the Wyoming Assessment Task Force. The Advisory Committee's major reasons for this recommendation are:

1. The quadruple counting of the college readiness test in the accountability system,
2. The lack of validity evidence for the ACT or SAT as an achievement measure,
3. The lack of independently verified alignment between the college readiness test and the Wyoming content standards,
4. The mismatch between Wyoming's allowable accommodations for students with identified disabilities and English learners and the accommodations permitted on both the ACT and SAT, and
5. The troublesome technical behavior of the achievement scores across years (or wobble in less technical terms).
6. The lack of instructionally-actionable information from the ACT score reports and associated materials.

We briefly expand on each of these points below. Further, the Advisory Committee recently became aware of an email memo (see Appendix A) provided to members of the Joint Education and Select Committees detailing ACT's response to public comments made at the September 21<sup>st</sup> Select Committee meeting. While the public comments were not made by members of the Advisory Committee or consultants, the ACT response put forth several assertions that the Advisory Committee felt were important to address.

---

<sup>1</sup> The brief was drafted by Scott Marion, Center for Assessment, but received a consensus recommendation from the Advisory Committee to forward it to the Select Committee for Statewide Education Accountability.

### “Quadruple Counting”

The 2015 Assessment Task Force and the Advisory Committee support the use of the college readiness assessment as an indicator of college readiness, but they are concerned that the ACT (the current college-readiness assessment) is used for achievement, growth, tested readiness, and as a major component of Hathaway readiness. The Wyoming educational leaders on both committees take issue with using the same score for four different things. The Advisory Committee welcomes the use of a college-readiness measure as one indicator of college readiness, but strongly objects to its use as an achievement measure for reasons we articulate throughout this memo.

### Validity

Tests are validated for specific purposes. The existence of evidence and logic to support claims of validity for one specific purpose generally does not make a test valid for other purposes. This is similar to saying that if a certain medication is proven successful to treat high cholesterol; it does not mean that it is also useful for treating high blood pressure. It might be, but we would want evidence of effectiveness for the new purpose.

College readiness assessment providers focus their evidence collection on predictive validity, which evaluates how well test scores predict some subsequent outcome of interest. In this case, both the ACT and SAT evidence is primarily focused on the relationship between test scores and students’ first-semester grades or first-year grade point average (GPA) in college, performance in college-level courses, and other college-level outcomes (e.g., retention and graduation). In other words, the use of ACT and SAT scores for predicting first year grades is supported by reasonable evidence. On the other hand, the Committee is aware of only limited evidence necessary to support the use of these assessments as indicators of high school achievement. The evidence necessary to support their use as achievement indicators is quite extensive, but we address only alignment and test consequences (below).

Chris Kratzer, the ACT representative, claimed that “the ACT is and has always been a curriculum-based achievement test” by virtue of their national curriculum survey. We appreciate ACT’s effort to collect data on the most common content elements and skills being taught in our nation’s classrooms. However, the United States has over 14,000 school districts, most of which have local control over their curriculum, so we are not clear how such a survey makes the ACT a valid achievement test as a state test for Wyoming. Such survey results are at best insufficient to support the use of the ACT as an achievement test and at worst, irrelevant. Further, as we describe in subsequent sections, a test used for achievement under ESSA must be able to report credible scores according to standards-based performance levels, which ACT seems unable to do.

## Alignment

Alignment is a foundation of standards-based education. It serves as a promissory note that says to educators, “we (the state/district/etc.) expect you to teach the knowledge and skills represented by these content standards, following which we will assess your students to see how well they have learned these standards.” Alignment is a measure of the degree to which a test fairly and accurately represents the set of knowledge and skills called for in the standards. We describe issues of alignment for both ACT and SAT below.

The major concern for using ACT as an achievement measure is that it is not designed to measure the content standards the Wyoming State Board of Education tells teachers it expects them to teach. Such a mixed message breeds incoherence when we are trying as hard as possible to create a coherent assessment and accountability system. This is not just hearsay from concerned members of the Advisory Committee, but it is corroborated by ACT’s response to the 2016 Fordham and HumRRO studies evaluating the quality of PARCC, Smarter Balanced, ACT Aspire, and MCAS:

*The finding that ACT Aspire assessments adequately assess many but not all of the priority content reflected in the Common Core standards is not surprising. Unlike other assessments included in the study, ACT Aspire is not and was **never intended to measure all of the CCSS**. Rather, ACT Aspire is designed to measure the skills and knowledge most important in preparing students for college and career readiness. This is a significant philosophical and design difference between ACT Aspire and other next generation assessments. ACT has made the choices we have to align with college and career readiness standards, rather than specifically to the Common Core, **and we intend to keep it that way**.*

While the study did not evaluate the quality of ACT, the relationship between ACT and the Common Core is the same as the relationship between ACT Aspire and the Common Core. If ACT is the state’s college admissions assessment in grade 11 and is used as an achievement indicator for accountability, the State will introduce confusion into Wyoming high schools. The question for educators will be whether they should focus instruction on the officially adopted state standards or on the ACT test specifications.

Chris Kratzer stated in his ACT memo that “the ACT is highly aligned to Wyoming’s state standards.” There is a tight relationship between the Wyoming content standards and the Common Core State Standards, so given ACT’s view expressed in the quote above, it is hard to understand how Mr. Kratzer can make such a statement. Further, it is one thing to assert that a test is aligned to certain standards and it is another thing to provide evidence to support such claims, particularly evidence drawn from independent alignment studies. The Committee has not seen any such evidence, but strongly believes that if evidence existed, it would have been shared with the Wyoming Department of Education and all other state clients.

Similarly, the new SAT is purportedly aligned with the Common Core State Standards, but to date, no independent evidence supports this claim.

### Allowable Accommodations

One of the most important advances of the No Child Left Behind era has been the expansion of opportunities for students who were either previously left out of the testing process or whose achievement was noticeably underrepresented. Many of these advances have come from the extensive research and development into test accommodations for students with disabilities and English learners. Accommodations are changes made to the standard test administration, presentation, or response conditions to provide students with fair opportunities to show what they know while not giving them an unfair advantage. States, consortia, testing contractors, and research organizations have worked to expand the types of allowable accommodations available for students as long as the accommodations do not provide an advantage for eligible students.

Both the ACT and SAT allow fewer and a narrower range of accommodations than essentially all state tests. On one hand this is understandable, because they need consistency across all states in order to serve as a comparable college readiness indicator. On the other hand, this policy is based on the voluntary use of these tests for college entrance purposes and not for a census academic achievement test for school and/or educator accountability. In fact, both the ACT and SAT have recently come under scrutiny over concerns that certain students were being denied legitimate educational opportunities. This plays out when students used to certain assessment accommodations on their state assessments are denied the use of the same accommodations on the ACT or SAT. These students (and parents) are faced with a choice: Taking the test using only the allowable ACT/SAT accommodations and potentially risking a lower score or taking the test with their regular accommodations but not receiving a “college reportable score.” In other words, to use the ACT/SAT for admissions purposes, the student may have to test again—and pay for it on her/his own—with the more limited accommodations. Again, this issue rears its head when the state requires all students to take the test, particularly for accountability purposes.

### Technical Issues

A key feature of standards-based educational systems is reporting the percentage of students scoring at various performance levels, established through carefully designed deliberative procedures that results in “cut scores” separating the performance levels (such as “proficient” or “basic”). In all current standards-based systems used for accountability, states must report the percentage of students scoring at the “proficient” level or better. Stakeholders hope to see an increasing proportion of students scoring at this level over time. When they see changes in proficiency rates, they should be able to

attribute these changes in test scores to true changes in student achievement and not to some fluke in the score scale.

Unfortunately, this is not how the ACT has worked in Wyoming. Figure 1 below displays the number of students scoring at each point along the ACT scale in reading for 2010-2011, 2011-2012, and 2012-2013. The fluctuations in the number of students at each score point are striking. We highlight just one example. In 2010-2011, 487 scored 18, an important college readiness benchmark, but just one year later, only 237 students scored 18, a drop of approximately 4.7% (based on an average number of 5300 test takers). This is a considerable problem at the state level, but even more egregious at school and district levels where fluctuations of 20% or more were common.

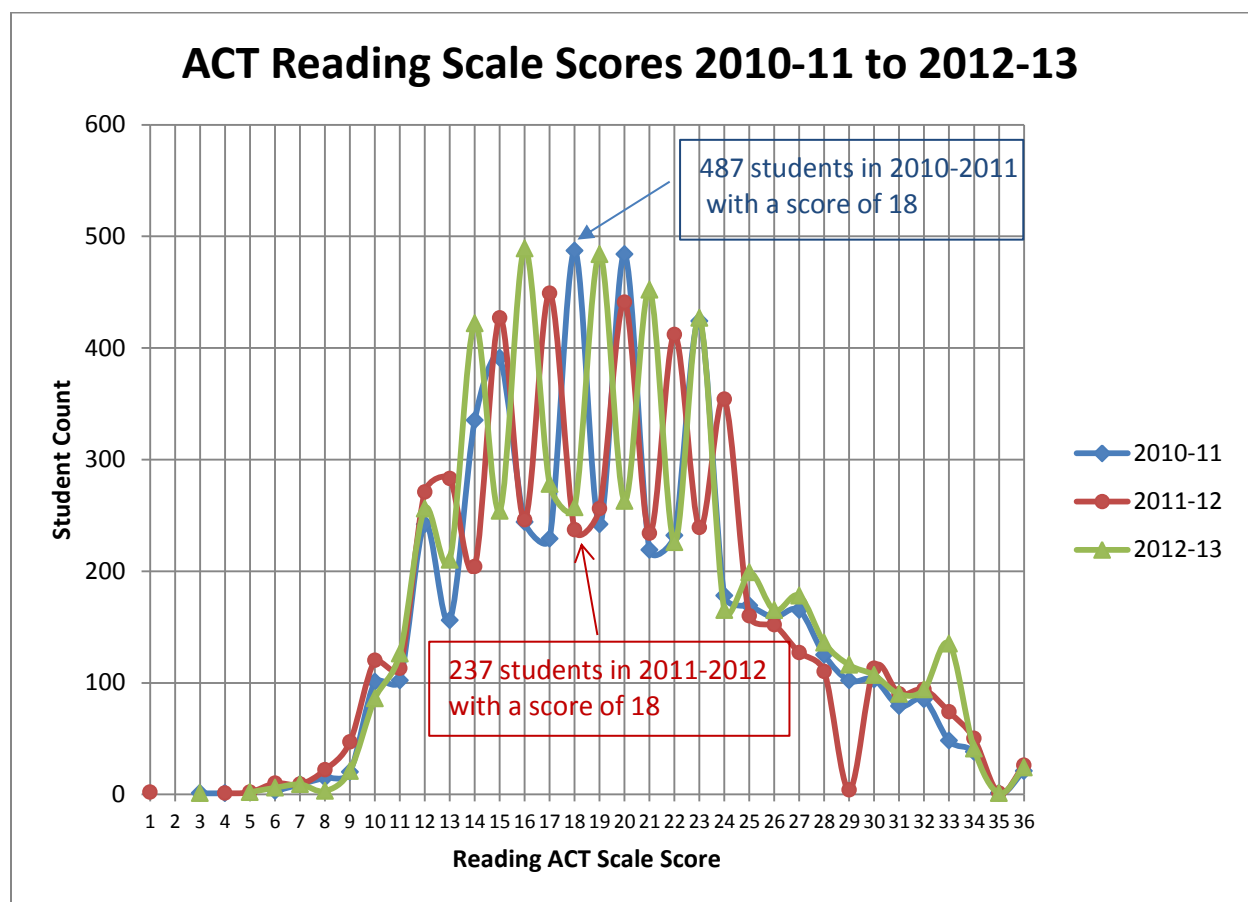


Figure 1. Wyoming ACT reading scores by year (2011-2013).

When Figure 1 was shown to Drs. Bob Brennan and Michael Kolen, two of the world’s leading psychometricians and the inventors of the ACT scale, their immediate reaction was “oh, they are just using a single form.” In other words, the ACT scale is designed to

be stable over multiple forms, but when one form is used as is the case with a census administration for a state, this kind of pattern can result.

Perhaps such a pattern can be excused in the first few years of an assessment program, although we do not think so, but upon examination of the most recent results, we are surprised to see that the pattern has not improved at all. Figure 2 below portray the ACT Reading results for 2014-2016, the most recent three years. As seen in this figure, the pattern has not improved at all. Again, this illustrates the concerns with using a test for a purpose for which it has not been validated. In fact, the Advisory Committee finds it inexcusable that the ACT claims to be an “achievement test” as noted in Mr. Kratzer’s memo, yet it cannot meet the most fundamental reporting requirement—reporting the percentage of students scoring proficient—for a standards-based assessment.

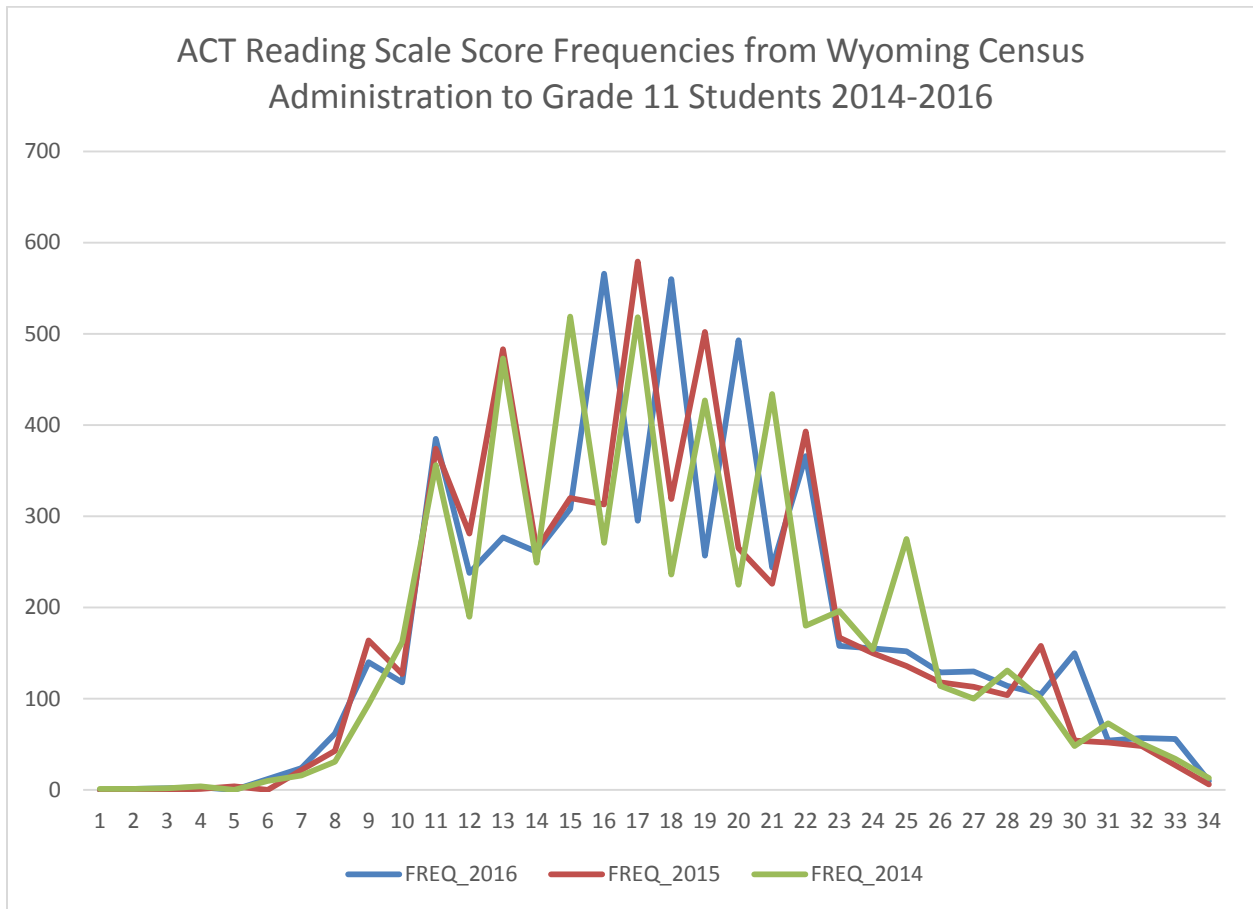


Figure 2. Wyoming ACT reading scores by year (2014-2016).

### The lack of instructionally-useful information from the ACT score reports

Many members of the Advisory Committee have direct experience administering the ACT and receiving the score reports. Compared to the rich information available from PAWS and other assessments, Advisory Committee members were disappointed in the granularity of the information received from ACT. In examining the score reports from ACT, it appears that the reported information is so general that educators have little information on which to base instructional actions to help students improve or more realistically how district and school personnel can use the information to improve curriculum, instruction, and interventions.

### Summary and Discussion

The Advisory Committee to the Select Committee on Statewide Education Accountability wrote this memo in an attempt to clearly and concisely outline the advisory committee's concerns with using Wyoming's college readiness assessment as a measure of achievement of the state's academic content standards. The concerns expressed in this document build off of those expressed by the Wyoming Assessment Task Force in its 2015 report to the Wyoming legislature, but expands in several important areas. We argue that this document presents a clear case against using the two widely-accepted college readiness tests, ACT and SAT, *as a statewide achievement measure* as part of Wyoming's school accountability system. The Advisory Committee acknowledges that it supported using the ACT for accountability when it was first proposed, but that was based on the assumption that using the ACT for multiple indicators was a temporary fix until a new assessment system could be designed and procured.

On the other hand, the advisory committee supports the use of a college readiness test *as part of the postsecondary readiness indicator* in the school accountability; although the committee has concerns about double counting the test results in this indicator as part of the school accountability system. The use of the college readiness test as part of the growth indicator—as the terminal test—is subject to many of the concerns discussed throughout this memo, but the advisory committee believes that the cost of using the college readiness assessment in growth calculations is outweighed by the benefit gained by having a growth measure that continues through 11<sup>th</sup> grade. This belief is driven by the fact that the methodology used to calculate growth scores can ameliorate some of the concerns to some degree.

## Appendix A: Memo from ACT to several Wyoming policy leaders

To \_\_\_\_\_

At a recent meeting of the Joint Education Committee on September 21, there were comments made about the ACT in the course of testimony that were not accurate. ACT knows that that the Wyoming Department of Education, Joint Education Committee members, and all other stakeholders want policy decisions about Wyoming's assessments to be based on accurate information. Accordingly, I am writing today to provide the following information in order to clear up any confusion that may have been caused as a result of these comments.

The statements in need of correction from the September 21 meeting are *underlined in Italics* below.

*The ACT is not a placement test:*

**The ACT is utilized throughout the United States to accurately place students in postsecondary courses that are commensurate with students' foundational skills.**

Postsecondary institutions throughout the country utilize ACT's College Readiness Benchmarks to determine the courses for which a student has the foundational skills and knowledge to be successful. The ACT College Readiness Benchmarks are scores on the ACT subject-area 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 college courses include English composition, college algebra, introductory social science courses, and biology. Based on a sample of 214 institutions and more than 230,000 students from across the United States, the Benchmarks are median course placement values for these institutions and as such represent a typical set of expectations.

Beyond our nationally-derived benchmarks, ACT routinely partners with states to determine state-specific benchmarks based on the state-specific postsecondary outcomes, and we are happy to do so in Wyoming.



The ACT is not aligned to Wyoming's state standards:

**The ACT is highly aligned to Wyoming's state standards.**

Wyoming, as well as a number of other states, is using the ACT statewide for federal accountability purposes and has submitted documentation and evidence through the peer review process. The peer review guidance requires that the assessments must be aligned with the “full range of the State’s academic content standards.” Should the peer review process reveal gaps, ACT has pledged to work with states to augment the ACT test so that it will more fully align to a state’s content standards in order to ensure an assessment that satisfies ESSA requirements and provides access and opportunity for students to every college and university in the nation.

The ACT is not an achievement test:

**The ACT is and has always been a curriculum-based achievement test. It is the principle on which our organization was founded.**

Every three to four years, ACT conducts the National Curriculum Survey which, in part, collects data about what elementary, middle school, and high school teachers do and do not teach their students. This data helps to inform what entering college students should know and be able to do to be ready for college-level coursework in English, math, reading, and science. The results of the survey are used for many purposes, but one of the most important is to ensure that ACT’s curriculum-based assessments—the ACT and ACT Aspire—continue to reflect what is occurring in the classroom in the context of college and career readiness.

ACT values our longstanding service to and partnership with the state of Wyoming, so please let us know if there is any additional information that would be helpful, or if there is anything that we can do to assist with your efforts to improve Wyoming’s assessment system and educational outcomes.

Sincerely,

Chris Kratzer

Senior Director

ACT, Inc.