

PARCC versus ACT/ASPIRE - The Future of High-Stakes Testing in Illinois

Illinois has committed to the PARCC consortium for its future assessments for federal accountability. Illinois also has a long-standing relationship with ACT, which is preparing a suite of competing third through tenth grade tests to compete with PARCC, as well as revising its ACT. There also is a second multi-state consortium, Smarter Balanced, which also is developing a set of assessments. Given the delays and uncertainty of PARCC's ability to deliver the promised assessments, the unknown quality of those tests, and the temporal and financial demands of those tests, many have questioned whether we should not continue with ACT products, some of which already have an acceptable track record. This paper explores the pros and cons of these options.

ACT, Its Track Record, and the Potential for ASPIRE

The ACT test is almost universally administered to Illinois students and has a decades-long track record of reliability and validity for its composite score, English test, and mathematics test. While score findings from its reading and science tests are not very meaningful (ACT has both indirectly and directly acknowledged the inadequacies of the reading test), the ACT English, math, and composite scores can provide an accurate picture of how well our students are learning in general. The test has a much higher level of importance to high school students than most standardized tests, since higher levels of performance will significantly enhance student options for college choice and financial assistance.

ACT also has the College Readiness Standards. These are a well-written set of standards that can be used to guide curriculum efforts. The degree to which they tie to the subject tests varies, but the standards for all subject areas describe skill and content knowledge that are needed by high school students. The standards have some commonality with the Common Core standards (not surprisingly, since ACT had a hand in developing the latter).

Hence, ACT has known value, albeit with limitations. It should be seen as a language arts and quantitative test (just as the SAT used to be packaged), given the reading and science test limitations. For a summative test designed to be used as an aid to making decisions about college admissions, the testing of those two areas should be adequate. Separate science and reading tests may be more feasible at lower grade levels.

The lack of a very valid reading test is ameliorated by the fact that reading skills are measured in the other ACT subject tests (and on most any standardized test). But in current practice, it is the ACT reading test that is used specifically to measure reading for NCLB assessment at the high school level. Such heightened importance is very inconsistent with the test's limitations

The problem appears to be that, particularly at higher grade levels, reading tests incorporate interpretation, which can become subjective. Increased subjectivity reduces reliability and therefore validity relative to more objective tests. Such subjectivity may be the reason the ACT reading test (as well as the PLAN and EXPLORE versions of the same test) has proven to be a poor predictor of high school or college classroom performance.

The ACT reading test therefore is very inappropriate for future use as a primary indicator of student performance. ACT would have to develop a more reliable language arts measure, possibly one more focused on its distinctly-superior English test, to have an acceptable language arts test.

Beyond the ACT tests, ACT products are a bit more questionable. The current PLAN and EXPLORE, while reasonably similar to the quality of an ACT, do not fit well into a high school schedule. The EXPLORE has been okayed for use by ACT from the beginning of 8th grade to the end of 9th and the PLAN from late 9th grade until the end of 10th grade. This arrangement works when a school is matching a test to the capabilities of its students, since student performance varies greatly among schools. But for use as a

grade standard, the EXPLORE is unsuitably lacking in rigor for use beyond the beginning of 9th grade. The PLAN is similarly unsuitable beyond early 10th grade. Additionally, its largely pre-algebra focus makes the EXPLORE's math test unsuitable for the increasingly high standards being set for math.

One current fix has been to issue a retired ACT for the end of 10th grade. But the ACT is not fully appropriate for 10th grade, since it has content not necessarily encountered until 11th grade. Clearly, an additional test is needed, so that there is one ACT test for each of the grades 8 through 11.

An additional and significant problem is the scaling of the ACT, which stems from the relatively small number on items on the test. The growth from one test to another across years often is only one or two points. One off-content or less than stellar item can change a student's test score, and hence growth, by a point. So it is not possible to accurately determine from such scores whether a student is progressing appropriately.

Finally, in terms of diagnostics, ACT links scores with its College Readiness Standards, which are very useful. But ACT does not provide additional documentation with its tests to connect test items with concepts. We cannot tie the items to the College Readiness Standards or identify curricular gaps by looking at clusters of items. If a school desires such information, it must conduct such analyses on its own, and such efforts, even if economically feasible, may not be sufficiently accurate.

The EXPLORE and PLAN will be replaced by the ACT ASPIRE system, due to be launched in Spring 2014, though conversations with ACT would make such a schedule seem optimistic. It is a suite of formative and summative tests for third through tenth grade. The ACT will remain largely the same for now, although it will be able to be computer-administered by 2015 and may include "Optional constructed-response questions."

The rigor of the test to be given in 9th and 10th grades will need to be higher than the current EXPLORE and PLAN, particularly for the math test of the former. However, only one test will be used for both grades, and the test is not a computer adaptive test, which has the potential to cover wider performance ranges.

The test will be longer, presumably due to more items, as seen at the chart on the right. The English test length stays the same for all tests, but all other tests increase in length. So the ASPIRE is 37.5% longer than the EXPLORE and 43.5% longer than the PLAN. The longer test is problematic in terms of scheduling but necessary in terms of testing psychometrics - the longer the test, the more likely it will be that the test will be more reliable and therefore valid. Incidentally, for comparisons to PARCC, the 30 minute ASPIRE writing should be added.

Test Length Contrasts for EXPLORE PLAN, and ASPIRE
(in minutes)

	EXPLORE	PLAN	ASPIRE
English	30	30	30
Math	30	40	50
Reading	30	20	45
Science	30	25	40
Composite	120	115	165

However, the additional length is a function of other needs. A 2012 survey showed that over 70% of the suburban schools gave the EXPLORE to 8th graders. For higher grades, a longer test is more appropriate, since older children can handle longer periods of testing. More importantly, the change from a test for each grade to one covering two grades requires a wider range of item difficulty levels, requiring significantly greater length. So the additional length may not have much impact on increased reliability or the ability to differentiate between grade level-based performance.

The current scaling of the EXPLORE, PLAN, and ACT (based on a 1 to 36 scale) has been inadequate to track student growth across years and diagnose whether a student is growing appropriately. Such a diagnostic function is of critical importance. ASPIRE will offer a three-digit scale. This would be superior

to the 1 to 36 scale, but only if it is based upon sufficient items to make the score differences truly meaningful. As discussed above, the individual tests remain relatively brief (particularly the potentially useful English, which with the math is the most reliable and valid EPAS test). ACT's desire to claim to have a reading and science test (tests which are by far their weakest ones) within a test of a relatively brief length limits their ability to have the longer language arts and math tests needed for tracking student growth. **So at the high school level, given the need to cover two grade levels, there is a likelihood that the ASPIRE subject tests are not long enough to provide a sufficiently greater ability to track growth from year to year.**

The ASPIRE will be much more expensive than the current EXPLORE and PLAN. ACT has proposed a cost of \$24 for the four subjects (\$25 with writing). The current cost for an EXPLORE and PLAN (no writing test) are \$10.50 and \$12.95, respectively. So the increase from EXPLORE to the Computer Based ACT Aspire, both without writing, is 128.6%. However, in terms of comparable paper-to-paper versions, there is a 185.7% increase, since ACT is adding a \$6 surcharge for paper. Since there is only one ASPIRE test, the final cost for the PLAN replacement is the same, \$24/\$30. So the increase from PLAN will be: 85.3% for the computerized version, and 131.7% for the paper. This may mean that there will be much room for negotiating for states adopting the ASPIRE, since preliminary information shows limited differences between the former and current tests.

Incidentally, the comparability of the paper-and-pencil tests to computerized tests is an issue for the ASPIRE (and PARCC). We do not know how comparable they really will be, and the differences may be significant.

One of the new wrinkles with ASPIRE is the use of constructed response items (which, as mentioned above, are planned for ACT). Such short-answer responses can demand higher levels of thought and be more of an authentic measure of student knowledge and skills. However, they also are more subjective, requiring well-trained education professionals to grade them accurately (not relatively low-paid staff with rubrics or initial attempts with computerized grading). Otherwise, the reliability of such tests will be low, and scores will be of little value. AP tests, which utilize subjective items, cost around \$90 and do have education professionals grading them. However, it is unlikely that ACT's costs will allow this. Additionally, one colleague reported that a proposed constructed-response ASPIRE item was created by dropping the multiple choice responses from a PLAN item. While this may work, it does not demonstrate that the tests have been significantly upgraded.

ACT speaks of tying the ASPIRE scores to its standards, which ACT already does with EPAS. Whether they give needed additional item cluster detail is uncertain. We need such information to diagnose our students' needs, and ACT's past resistance to such requests has not been encouraging.

Also, ACT describes the ASPIRE as entrenching the College Readiness Benchmarks into its measures. The predictions of college readiness from the benchmarks bear little resemblance to actual college performance. Such benchmarking will keep schools on the defensive as reports of low levels of college readiness continue to be reported to the media annually. The low levels of benchmark readiness are due, in part, to the fact that the individual tests predict a limited amount of college course performance, particularly with the reading and science tests.

The disconnect between the benchmarks and actual college performance also is due to the inherent lack of logic in the presumptions underlying the presentation of the benchmarks. Students can succeed very well in college without being strong in all subject areas. Students almost never major in subject areas in which they are weak. Such entrenching of flawed readiness data in the test feedback report will necessitate a dismissal of part of a report that we are trying to get the student to take seriously. We already have this problem with EPAS score reports, and the new system may make the problem much worse.

While the ACT has some alignment to the Common Core Standards, it is not nearly as aligned as it could be. ACT's alignment is self-reported, not independently verified. Since the 9th and 10th grade ASPIRE tests seem to be not that different from their predecessors, they probably will be similarly semi-aligned. We have no idea of whether the other ASPIRE tests will be aligned.

The imminent release of ASPIRE brings up another concern with ACT - their other, newer products have, at best, not distinguished themselves. In ACT's recent workshop and in discussions with their sales staff, it is very apparent that ASPIRE is nowhere near ready for administration, and the rush to a Spring 2014 release is likely to sacrifice test quality and freedom from administrative glitches. Their computerized community college placement test, the COMPASS, was recently dismissed as having very limited value by two 2012 studies from the foremost community college research department in the country, the Community College Research Center at Columbia University. Their popular WorkKeys test never has received any validation from a scientific research study. Their Quality Core, an attempt at packaging curriculum and assessments, has thus far received no positive feedback in Chicagoland of which I am aware. I have seen no independent reviews on their other products. Finally, their proposed K-2 assessments are highly suspect in terms of potential validity or value – the proposal seems to be an action borne out of opportunity rather than genuine need. So in terms of trusting ACT beyond their original test and its direct offshoots, the evidence is not positive.

There are other issues with the current use of ACT. ACT has not allowed states to use the ACT as their state test for NCLB without combining the ACT with the never-scientifically validated WorkKeys test. ACT's flexibility on adapting to state testing needs are limited by their primary focus on remaining consistent to the restrictions they consider necessary to producing a national college entrance examination. Such restrictions have to be resolved with ACT's emerging role as a provider of an annual test to a universal audience within a state.

Additionally, ASPIRE administration is problematic for a wide array of schools across the country due to its focus on computer-based administration, causing one state so far to withdraw on that account alone. ASPIRE can be taken as a pencil and paper test, but that will cost six dollars more per student. Yet, as mentioned earlier, it has not proposed to create a computer-adaptive test, one of the main reasons for using a computer-based test. Computer-adaptive tests have the potential to more accurately measure student performance by tailoring the items to the student's ability level (though the effort to precisely identify item difficulty, essential for such a test to be truly valid, requires a substantial investment of skilled educator and psychometrician time). Such tests can provide a greater level of vertical integration, enhancing the potential to more accurately measure progress across time. ACT's focus on computer-administered tests without the logical extension of using computer-adaptive format is very questionable. Please note that ACT's effort at computer-adaptive testing, the COMPASS, has been of very limited validity itself.

Furthermore, many schools are entrenched in Mac-based computing, particularly with the growing dependence on IPADS. ACT has largely ignored the Mac world with its computer-based products (e.g., the COMPASS cannot run properly on Macs even with the integration of a virtual PC platform). ACT has claimed that this will change. In 2015, the ACT products are supposed to work on IPADS as well as android tablets.

Finally ACT has shown that, while it technically is a non-profit institution, it has functioned like a market driven for-profit entity. It has heavily marketed its products for years, ignoring research and even common sense that would warrant much more restraint. In short, ACT has been trustworthy for only parts of a test which it had developed and refined decades ago, and for the tests directly derived from the good ACT tests. Even these good tests have been too short to diagnose student growth, and the ASPIRE does not seem to

resolve such issues. Its other products are suspect, as are its motivations for promoting its products and services. We cannot assume that they will develop a sufficiently useful suite of tests for the future.

PARCC

PARCC has been presented as the next generation of testing, with a regional focus so that each state would not be reinventing the wheel. Such pooling of resources promised to produce a test that would be aligned to the new common core standards and superior to what any single state could produce. Its proposed greater length promised greater reliability and hence validity, its performance tests promised greater authenticity, and its differentiated formative and summative administrations promised a more frequent use of its data and potentially greater integration into classroom teaching.

But it has not been able to meet its originally proposed timelines. There is widespread concern that the consortium will have further defections as PARCC's delays undermine the trust of consortia members. Currently, if the consortium drops below a certain number of member states (reportedly 15, though I have not been able to verify that number), it will lose federal funding and possibly dissolve. At this writing date, New York and Massachusetts have just delayed the implementation of their accountability using PARCC, and Kentucky has withdrawn, so they are about at that minimum number.

PARCC's proposed length is a glaring administrative issue. Some of the current ACT, PLAN, and EXPLORE tests are too short to be sufficiently reliable for a next generation test. But schools only have so much time to teach students, and tests are not necessarily the best use of instructional time. The expected multiple administrations of the test (both formative and summative) exacerbates the problem. PARCC states that "the estimated time it will take students to complete all of the sessions of the performance-based and end-of-year components in both ELA/literacy and mathematics at each grade level" will be from 8 to 9 hours in grade school and 9.5 hours in middle and high school (ISBE reports that the time required may even be more than 9.5 hours). Additionally, PARCC has stated that for the test, we will need to allocate an additional 50% to the 9.5 hours for any student needing more time. Furthermore, PARCC has required that extended-time students (not just students wanting more time) are allowed as much as a full day for each testing session, to the point of having a certified staff member accompany them to lunch. They also have stated that only certified staff can coordinate and administer the tests. All of these restrictions will create immense burdens for many schools.

Not counting classroom related needs of distributing the tests, collecting the tests, and allowing for breaks, we currently require 4.5 hours for ISAT language arts and math testing for grades 3 through 8 (science testing in 4th and 7th grades adds an additional 1.5 hours). At the 11th grade, we require 6 hours and 20 minutes of PSAE testing. The desirability of a longer and therefore potentially more accurate test like PARCC must be weighed against its intrusion into instructional time.

There are other major time-related issues. At the high school level, most students will need to continue to take the ACT. Before universal testing, 70% of Illinois high school students took the ACT, and that percentage probably will be higher now. The advantage of ACT is that potentially college-bound students must take it seriously. For PARCC to command similar respect from test-takers, it must earn the respect of colleges to the extent that they will use it in place of an ACT or SAT. Colleges cannot do that without first validating the tests, although only 640 (less than one-sixth of the United States' approximately 4,000 higher education institutions) have "committed to participate in PARCC" (according to PARCC). So we can anticipate using both ACT and PARCC for at least two or three years until (and if) higher education decides that PARCC tests are sufficiently reliable and valid measures of college readiness. Given the much greater temporal requirements of PARCC, such parallel testing will be particularly demanding.

PARCC does not see its assessment as becoming a college admissions test. "The assessments are not intended to be used by colleges and universities in decisions about college admission. They are intended

only to give students a College- and Career-Ready Determination that indicates they are academically prepared to enroll in first-year, credit-bearing courses at two- and four-year institutions and can be exempt from taking a placement test at the college or university they attend." But if it is not used for such a purpose, it will not be taken very seriously by most students.

The current timing of their administration window overlaps with spring break and, at the high school level, AP testing. This will be a significant problem for high schools, particularly those with large AP programs.

PARCC tests will not include science. But then, the ACT EPAS tests never included a meaningful measure of science, either, and the state's current PSAE science test has received minimal attention. At the high school level, there may not be any way to measure science well, due to the differences in courses taken at a given grade level (if at all) - it makes little sense to give a test with many chemistry items to a freshman who has not yet had chemistry. Also, creating a relatively short test to cover all of the central concepts in biology, chemistry, physics, earth sciences, and other subject areas is infeasible - it would take hours.

As with ASPIRE, PARCC administration is problematic for a wide array of schools across the country due to its focus on computer-based administration, causing one state so far to withdraw on that account alone. PARCC can be taken as a pencil and paper test, but that will cost \$3 to \$4 dollars more per student. And like ACT ASPIRE, it has not proposed to create a computer-adaptive test, again one of the main reasons for using a computer-based test.

PARCC has made some other questionable decisions. One that I personally find problematic is their way of determining whether a school is meeting standards. They have adopted ACT's benchmarking protocol; i.e., the standard for a test score will be set at the score point at which at least 75% of the students in a subsequent college class achieve a grade of C or better. Such benchmarking ignores the issue of whether the test is sufficiently accurate to be a reliable and valid predictor of student performance. In the case of ACT, such a benchmarking protocol was applied to tests with very little predictive value, and the same misuse can occur with PARCC (particularly since PARCC has had precious little time to refine their instruments).

The findings from PARCC's performance-based assessments are subject to caveats similar to those for ACT's constructed response items. They may be more valid in terms of their authenticity, but less valid due to their lower score reliability, since they are inherently more subjective than multiple choice items. A critical unknown is the issue of whom or what will grade the test and how accurate it will be. It is unlikely that there will be an ability to spend substantial money on grading by professionals in the field, and a computerized program or relatively low paid grader using a simple rubric may not provide valid measurement. Therefore, their ultimate value is unknown and questionable.

Finally, cost will be an issue. Preliminary examinations suggest that Illinois costs may increase substantially, though a small part of this may be offset by PARCC's inclusion of some of the processing costs now borne by the state. PARCC has proposed a cost of \$29.50 for its current test plan, but it may go up if more states drop out of the consortium. While the much greater PARCC test length provides justification for the approximately 20% higher cost, this alone may be used as a rationale by Illinois to go with ACT despite their current commitment to PARCC.

Other Issues

There remains the overarching issue of whether we really should attempt to test all core subject areas. Research has shown clearly that ACT only directly tests math and language arts well. Reading skills in specific are indirectly tested by all ACT tests, since such skills are integral to doing well on any ACT test. EPAS science tests may be useful only in that they lengthen the overall testing time and therefore may

increase the ability of the composite score to predict future classroom performance in all subject areas. The reading tests may also do this, although their subjectivity may be the reason that the reading scores sometimes reduces the predictive value of the composite scores (ACT excluded reading from its own benchmark validation prediction study). But both science and reading are relatively useless as science and reading tests, with very minimal connection to any subject area.

PARCC proposes to take a considerable amount of time to test math and language arts (the latter encompassing reading), but has not talked extensively about testing other areas. In general, a realistic approach to regional or national testing will focus on mathematics and language arts skills (e.g., the approach used for years by SAT). Science skills will be difficult to link to high school grade levels; e.g., will the 9th grade science test focus on biology or physics or what? Science at the high school level is too broad a topic to be tested without creating subject-specific tests, such as are offered by SAT and AP.

There has been occasional discussion of the possibility of PARCC merging with Smarter Balanced. Smarter Balanced has made three choices for its testing that seem very desirable. They have provided the option for districts to offer only the summative end-of-year tests, rather than interim, formative, and summative tests. At \$22.50 per test, it is the lowest cost option. Their tests are largely computer-adaptive, which, as mentioned above, are potentially superior. Furthermore they have opted to make all of their items available to the public, rather than struggle with the often daunting task of test security. By using an extremely large pool of test items, far too many for anyone to memorize for a test, and making them available for review, the need to secure test items is eliminated.

On the minus side, they do not have tests for 9th and 10th grades, a distinct liability. Also, I have found no provision for paper and pencil testing. Finally, their testing time requirement of from 6 to 7.5 hours, not counting accommodations, is a bit longer than our current state testing, although shorter than PARCC. However, they, like PARCC, have not put strict limits on time, so a student does not require special accommodations to be allowed extra time.

While we already have committed to PARCC, Smarter Balanced might be a useful option, particularly if the two consortia united their resources. Such pooling of resources, particularly if Smarter Balanced prevails on the computer-adaptive approach, may be the best option.

Additionally, there has been talk of College Board entering the competition. College Board has much more experience with subject-specific tests (with its AP and SAT Subject tests) and has been much more accurate and responsible in approaching assessment and appraisals of college readiness. They did not falsely claim to be able to test science, and only a few years ago succumbed to the political necessity of specifying a reading test. The verbal and quantitative tests were sufficiently long and comprehensive to be reliable measures. They also were much more accurate in their appraisal of college readiness, acknowledging a college readiness figure of 43% that is 72% higher than ACT's absurd 25%. Such a responsible approach, one that did not sensationalize college readiness, may have cost the SAT its market share, given ACT's aggressive market expansion in the past decade.

Conclusion

High-stakes testing truly is high-stakes. It requires many hours of student and teacher time. It provides students with often highly stressful hours of testing and labels them with a score rating that can have life-long implications. The aggregate scores can label schools as desirable or undesirable, impact teacher and administrator careers, and ultimately be used to judge the educational systems of an entire state or even country. As we stand on the threshold of a new testing era, we have to get it right.

The following bullet points summarize the issues related to the PARCC consortium.

PARCC pluses:

- They are focused on mathematics and language arts, the two most accurately testable areas.
- They recognize the need for longer subject-area tests for reliability and validity needs.
- PARCC tests were designed from the Common Core and should best represent them.

PARCC minuses:

- They have missed prior deadlines on their test roll-out schedule, and do not yet seem to be ready to deliver their tests.
- 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.
- A significant piece of the above time issue is their open-ended allowance for students needing additional time, which can extend their testing time to an additional 50% for any student wanting it.
- The current timing of their administration window overlaps with spring break and, at the high school level, AP testing.
- Students currently have minimal motivation to perform well on the tests.
- Most high school students will continue to need to take the ACT (3 hours and 25 minutes of additional testing time, excluding mandatory breaks) until/if the PARCC test can be used for college admissions.

We have serious concerns with PARCC, but it has the potential to be much better than what we have now. Despite sample items that look promising, PARCC quality is an unknown, with no track record. PARCC will take far more time than currently available testing products and will cost more money. It will require more of a revamping of our longitudinal school improvement tracking than if we continued to use ACT products. It will require eventual acceptance by higher education institutions in order for it to be meaningful to our students.

But we should not yet give up on PARCC. There remains much promise to the consortium effort and the fresh start it could provide to high-stakes assessment. The foremost needed modification would be to reduce the time demands. Modifying the test (particularly by lessening the time demands) so that it could fit into our current end-of-April time frame would make it much more acceptable to schools, and should lower the costs.

Continuing with our current ACT testing company's products has been considered as a viable alternative to PARCC. The following bullet points summarize the issues with ACT.

ACT Pluses:

- 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 the test takers.

ACT Minuses:

- ACT science and reading tests are weak predictors of future classroom performance.
- The degree to which the ACT products align to the common core is uncertain.
- The single ASPIRE test that is replacing both the PLAN and EXPLORE may be too short to track student progress accurately at the high school level.
- The ASPIRE test that replaces the EXPLORE and PLAN is much more expensive with little justification for such a large increase.
- Grade 3 through 8 ASPIRE tests currently remain an unknown.
- ACT does not yet seem to be ready to deliver the ASPIRE test.

- In the past, ACT has required Illinois to administer the never-scientifically-validated WorkKeys suite of tests.
- Newer products from ACT have been disappointing.
- ACT's corporate ethics have been questionable.

The ACT test has proven useful but needs to be significantly better. ASPIRE does not appear to be sufficiently different in content from the EXPLORE and PLAN to fix similar issues with them at the high school level (same subject tests, with similar inadequate length issues), or to justify the cost increase. Whether it will meet the 3rd through 8th grade needs is as unknown as is PARCC. Will we have true vertical integration despite the lack of a computer-adaptive format? Will we be able to identify growth? Will we be able to make connections between items, concepts, and standards? Given ACT's track record with anything outside of their ACT test and its EXPLORE and PLAN permutations, ACT's proposed future products are unknowns that cannot be assumed to be valid, reliable, or useful. And the ACT, EXPLORE, and PLAN themselves have needed significant improvement for years.

Finally, ACT has engaged in public education bashing for several years, despite substantial improvements in Illinois ACT performance (which ACT never mentions). Assuming that ACT will continue to mislabel us as failing schools who cannot prepare our students for college, we should not continue to support them.

Nevertheless, we should further explore the prospect of using ACT, simply to protect ourselves from a PARCC consortium collapse if too many states withdraw, or at least as a possible interim test until we have a more proven and administratively feasible next-generation test. We need to make ACT accountable to us if we do continue to use them, we need this discussion before they become what may seem to be the only choice, and we should not get into any long-term contracts.

Finally, this is not necessarily a choice between just PARCC and ACT. Smarter Balanced may be an option which should be investigated, though their lack of 9th and 10th grade tests could be a deal-breaker. With all of these uncertainties, it might be desirable, if possible, to continue our current testing for at least another year and not change until our potential vendors have had time to work out the bugs and present us with more finalized products. And while schools need annual testing, ISBE accountability testing could be limited to important transitional grades (e.g., 5th, 8th, and end of 11th), saving considerable time and money.

Moving ahead with PARCC should be predicated on three conditions:

- The length of the test will be reduced by about half (or more) of its current length. We additionally need to reduce the 50% additional time for those wanting it. One possible path to these ends would be to make optional the potentially more authentic, but also potentially less reliable, performance tests. Such moves also would cut the cost.
- It also needs to be scheduled in the school calendar near the current state testing time window. PARCC cannot interfere with AP testing in high school, and its proximity to spring break must be corrected.
- Higher education, at least in terms of Illinois institutions, will commit to using the PARCC for admissions if the test pilots as a reliable and valid instrument. If the PARCC test is designed to measure college readiness, and if research on an initial college freshman cohort validates this, then higher education needs to commit to honoring it as sufficient for college entrance and placement. Please note that PARCC currently specifies that it is not to be used as a college admissions test.

No matter which vendor serves us, we need to function as educated buyers of services, not captive clientele, and clearly push them to meet our expectations and needs. No one test or suite of tests can comprehensively measure students and schools. But tests can provide very valuable information. We need better tests, and now is the time to implement them.