Comparison of Electronic vs. Paper Assessments
Impacting Student Achievement and District Accountability in Ohio

Meeting with Senator Joe Schiavoni
March 15, 2016
8:45 a.m. to 9:45 a.m.

Prepared by: Matthew L. Bowen, Superintendent
Campbell City Schools

The purpose of this document is meant not to criticize electronic assessments or the Ohio Department of Education. The electronic testing platform assesses greater depth of student knowledge which encourages best teaching practices for all learners in Ohio. The document will address perceptions that have been created due to two different testing platforms administered to students in 2015, where a single cut score was selected to determine letter grade ratings. Parents, students and educators across the state may have a false reality of success and others a feeling of failure.

Mixed assessments, in one accountability rating system, are confusing to all stakeholders.

We absolutely agree that we must continue to integrate technology. All districts must responsibly build their infrastructure to support technology for daily instruction and assessments. All Districts must transition to electronic assessments where students’ knowledge will be assessed with greater depth and rigor.

All districts were to assess electronically in 2016. If this had been enforced, the need to advocate for change would not be necessary. Stakeholders would likely accept the confusion of accountability of the previous 2015 year and move forward. It is difficult for districts to move forward when the accountability system remains unchanged and two assessment platforms continue to be an acceptable practice.
Objectives:

To foster awareness of student performance and district accountability ratings, which may have been impacted by varied assessment platforms.

To advocate for public dismissal of 2015 testing results and future 2016 testing results. Data collected through multiple assessments, then combined, and a single cut score chosen was the previous practice in 2015 and the anticipated future practice for future testing. The single cut score is the basis for student performance levels, district ratings and teacher evaluations.

To advocate for all Districts to assess electronically in 2017 where a single accountability platform will allow for an accurate comparison of students achievement and growth.

Talking Points:

1. Describe similarities and differences between varied platforms used for student assessments (ie. electronic tests vs. paper/pencil tests)
2. Discuss the completion of the technology readiness survey prior to testing in 2015. Districts that responsibly built capacity with their technology infrastructure were required to test using the electronic platform.
3. Explain timing issue related to switching to a paper/pencil assessment in light of recent research. Discuss the lack of paper/pencil resources available by the State if all districts were to use the same assessment tool.
4. Discuss Michael Molnar’s research regarding the electronic testing and its relation to value added ratings; Michael is the Executive Director of Educational Services in Amherst Exempted Village.
5. Discuss impact of an electronic platform on testing scores of children who are assigned by USDE and ODE to certain subgroups.
6. Convey the gravity of the impact of Third Grade Reading Guarantee scores with respect to students who are economically disadvantaged.
7. Analyze assessment items from both platforms and discuss how questions differ with respect to depth and rigor.
8. Review high stakes testing and student retention in grade three. Employing the electronic assessments may adversely impact students identified economically disadvantaged.
9. Discuss how teacher evaluations may have been adversely affected by value added growth data. All student data collected from both platforms was combined then ranked.

Supporting Evidence:


“In just three days, 428 school districts across the state responded with data regarding their selection of testing format from last year. The highlights of the research findings are listed below.”
“Of the 428 school districts that responded as of March 4, 2016, 89 districts utilized the PARCC paper only assessments for all grades 4–8 math and ELA testing last year. As you can see, 85% of those districts received an A on the Overall/Value Added measure. In addition, 15 of the districts that received an F on the Overall Value Added measure in 2014 improved to an A in 2015.”

“Of the 428 school districts that responded as of March 4, 2016, 260 districts utilized the PARCC online assessments for all grades 4-8 math and ELA testing last year. As you can see, 62% of those districts received an F on the Overall Value Added measure. In contrast to the paper-only districts, only 3 of the districts that received an F on the Overall Value Added measure in 2014 improved to an A in 2015 while an astonishing 80 districts dropped from an A in 2014 to an F in 2015.”
Of the 455 school districts that responded as of March 7, 2016, 93 districts utilized the PARCC paper-only assessments for all grades 4 – 8 math and ELA testing last year. As you can see, 92% of those districts maintained or improved their Overall Value Added grade from 2014 to 2015. In addition, 41 of those districts maintained an A on the Overall Value Added measure and 38 districts improved to an A. Only 2 districts maintained an F."

*Of the 455 school districts that responded as of March 7, 2016, 291 districts utilized the PARCC online-only assessments for all grades 4 – 8 math and ELA testing last year. As you can see, only 41% of those districts maintained or improved their Overall Value Added grade from 2014 to 2015. In addition, 128 of those districts decreased to an F on the Overall Value Added measure with 71 of those districts plummeting from an A to an F. An additional 43 online-only districts*
maintained an F on the Overall Value Added measure compared to only 2 from paper-only districts.”

"Comparing Paper and Computer Testing: 7 Key Research Studies”  By Benjamin Herold (February 23, 2016)

“Earlier this month, officials from the multistate Partnership for the Assessment of Readiness for College and Careers acknowledged to Education Week that there were discrepancies in scores across different formats of its exams...with the advantage for paper-based test-takers appearing to be most pronounced in English/language arts and upper-grades math.”

“Briggs said computer- and paper-based versions of an exam shouldn’t necessarily be expected to measure the same abilities, or have comparable results.”

“A mounting body of evidence suggests that some students tend to do worse on computer-based versions of an exam, for reasons that have more to do with their familiarity with technology than with their academic knowledge and skills.”

“PARCC Scores Lower for Students Who Took Exams on Computers Discrepancy Raises Questions About Fairness”  By Benjamin Herold (February 3, 2016)

“Students who took the 2014-15 PARCC exams via computer tended to score lower than those who took the exams with paper and pencil—a revelation that prompts questions about the validity of the test results.”

“It is true that this [pattern exists] on average, but that doesn't mean it occurred in every state, school, and district on every one of the tests,” Jeffrey Nellhaus, PARCC’s chief of assessment, said in an interview.

“The pattern of lower scores for students who took PARCC exams by computer is the most pronounced in English/language arts and middle- and upper-grades math.”

“Several states have since dropped all or part of the PARCC exams, which are being given again this year.”

“Low- and middle-performing students did not similarly benefit from taking the exam on computers, raising concerns that computer-based exams might widen achievement gaps.”

“The State That Pulled The Plug On Computer Testing” by Blake Farmer (February 20, 2016)

“Even in recent weeks, system testing uncovered new problems.”

“All Tennessee students deserve a positive testing experience every time they log in, not one that is slow to load or fails periodically due to too many users or poor judgement on the part of the vendor.”

“Nearly two-dozen states have moved to online exams, many with the PARCC and Smarter Balanced consortia. And Scherich says many have run into trouble.”
“Responding to school boards around the state that have called for teachers and districts to be held harmless, Tennessee Governor Bill Haslam says this year's scores will only count toward a teacher's evaluation if they would help.”

Grade Three Test Questions (Paper/Pencil vs. Online):

<table>
<thead>
<tr>
<th>Paper Pencil</th>
<th>Online</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Which number line shows the correct location of the number ( \frac{2}{3} )?</td>
<td><img src="image" alt="Number line and point" /></td>
</tr>
<tr>
<td><img src="image" alt="Number lines" /></td>
<td>Select the place on the number line to plot the point.</td>
</tr>
<tr>
<td>11. Connie solved the math problem shown.</td>
<td>Bryan has 48 cupcakes in 6 boxes. Each box holds an equal number of cupcakes. Bryan uses this equation to find how many cupcakes are in each box.</td>
</tr>
<tr>
<td>[ 40 + 8 =? ]</td>
<td>[ 48 \div 6 =? ]</td>
</tr>
<tr>
<td>Which equation can Connie use to check her answer?</td>
<td>Create a different equation Bryan could use to find the number of cupcakes in each of the 6 boxes.</td>
</tr>
<tr>
<td><img src="image" alt="Equations" /></td>
<td>Select from the drop-down menus to correctly complete the equation.</td>
</tr>
<tr>
<td>10. Select the equation that is not correct.</td>
<td>Enter your answers in the boxes.</td>
</tr>
<tr>
<td><img src="image" alt="Equations" /></td>
<td>[ 9 \times 9 = ]</td>
</tr>
<tr>
<td>16. Select the three equations that are correct.</td>
<td>[ 63 \div 7 = ]</td>
</tr>
<tr>
<td><img src="image" alt="Equations" /></td>
<td>[ 30 + 5 = ]</td>
</tr>
</tbody>
</table>
17. Select the two shapes that have parts that are each $\frac{1}{6}$ of the area of the whole shape.

- [Image of shapes]

33. Eric measures 10 leaves with a ruler. He records the lengths as shown.

<table>
<thead>
<tr>
<th>Lengths of Leaves (inches)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5, 6.5, 6.5, 6, 5.5, 5.5, 5.5, 5.5, 6, 5.5</td>
</tr>
</tbody>
</table>

Which line plot shows the lengths of the leaves recorded correctly?

- [Image of line plots]

Use the More or Fewer buttons as many times as needed to divide the circle into 6 equal parts.

Then shade $\frac{1}{6}$ of the area of the circle. Divide the figure into the correct number of equal parts by using the More and Fewer buttons. Then shade by selecting the part or parts.
Conclusion:

A single cut score for each test was determined to identify the district’s ratings, though two assessment platforms were deployed to record the student performance. In school year 2014-2015 the ODE surveyed each district to determine how assessments would be administered, paper or electronic. Districts were not given choice and were required to test electronically if they were prepared. Through our initial review of state data and citations from national publications, children who participated using paper/pencil platform had a greater advantage to score higher than those students who participated using an electronic platform. This conclusion is supported, by data collected throughout Ohio, and also through research that identifies the trend nationally.

Considerations:

1. Our recommendation is NOT to abandon the electronic assessments, but rather forfeit the 2015 report card results, as well as future 2016 report card results during which students will still be assessed using varied assessment formats (paper/pencil and electronic). There will be greater confidence, reliability, and validity of results when there is consistency among assessment tools.
2. If the data remains unaltered and published for public review, the consideration suggested would be to segregate paper/pencil platform results and electronic platform results and assign two separate cut scores. By grouping and ranking students for proficiency, a more valid comparison can be made between student, district and teacher performance purposes. We understand that this may not be feasible, due to some districts using a hybrid approach of test administration, utilizing both assessment platforms.
3. We must re-consider the use of electronic testing in third grade. Student retention required by the Third Grade Reading Guarantee should not be applied to economically disadvantaged students who participate using an electronic platform.
4. Mandate all districts to assess electronically in 2017 where a single accountability platform will allow for an accurate comparison of students’ achievement and growth.