When studying school effectiveness, accountability metrics are no substitute for high-quality research conducted by independent researchers. Accountability metrics, typical of those used in Louisiana, are influenced by test score inflation and manipulations of the cut scores that determine passing rates. Researchers also warn of serious technical limitations of school proficiency rate metrics. Accountability metrics typically fail to account for the sizable influence that family background and school selection have on student achievement.

Researchers agree that it is better to focus on the actual test scores rather than group passing rates. Value-added and other quasi-experimental techniques use the actual test scores (or statistically appropriate equivalents) rather than passing rates and measure individual student achievement growth between two points in time, seeking to eliminate the impact on learning growth of student characteristics, family, peers and community—factors not controlled by teachers or schools. Growth in school proficiency rates is not a value-added measure because one group of students in one year is compared to a different group of students the next year. This fact and other technical shortcomings led the Albert Shanker Institute’s Matt Di Carlo to warn, “If your evidence is changes in proficiency rates, you probably don’t have much evidence.”

Several high-quality research studies suggest markedly different stories about the New Orleans experiment.

- **Pre-Katrina students in Recovery School District schools in 2009 showed no net growth after Hurricane Katrina.** Evacuee students, who returned to New Orleans and attended RSD schools in 2009 (the schools in the state takeover following Hurricane Katrina), experienced no statistically significant post-Katrina achievement gain measured by value added.  

- **The highest evacuee achievement gains were in traditional public schools outside New Orleans.** Through 2009, evacuee students from New Orleans who remained in traditional public schools outside New Orleans had much higher achievement gains than students returning to schools in New Orleans.
Evaluating the impact of New Orleans reforms is different from studying charter school effectiveness. Two high-quality research studies focused on comparing charter schools to non-charter schools within New Orleans during the years when non-charter schools (primarily the RSD direct-run schools) educated large numbers of students. Both found charter schools were more effective than RSD direct-run schools. This research focus, however, is largely a historical distraction now that the RSD is all charter and the RSD direct-run schools for the “leftover” students have been shut down. The authors of one of these studies points out that their charter school study, and others like it, did not estimate causal effects of the New Orleans portfolio (school choice) system. They said that would require data for the entire state of Louisiana producing value-added comparisons over time of all New Orleans schools to all other schools in the state. In other words, the academic impact of the post-Katrina reforms must be judged by the impact of the takeover on all students, including those in the direct-run RSD schools, not just those enrolled in charter schools.

Charter school effectiveness declined as charter school concentration increased. In 2011, according to the Center for Research on Education Outcomes’ 2015 urban charter school study, charter school effectiveness dropped by half in math compared with earlier years and dropped even more sharply in reading. The drop-off likely occurred because charter schools had grown in number and were enrolling more difficult-to-educate children from shuttered direct-run schools. The 2015 CREDO study did not include 2012, 2013 or 2014 data when nearly all students attended charter schools and when the common application system was in effect.
Learning Gains for Charter School Students Relative to Virtual Twins in the Control Group of Non-Charter Schools

<table>
<thead>
<tr>
<th>NCE Relative to Control Group</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.9</td>
<td>3.2</td>
<td>2.9</td>
<td>2.9</td>
<td>1.5</td>
</tr>
<tr>
<td>Reading</td>
<td>1.9</td>
<td>2.5</td>
<td>2.3</td>
<td>1.9</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Standard deviation units converted to NCEs by AFT. An NCE of 0.0 represents the control group’s average learning gain with approximately 21 NCEs equal to a standard deviation. NCEs are roughly comparable to percentiles.

- **Test scores changed modestly between 2008 and 2011.** According to a high-quality study by University of Southern California (USC) researchers, New Orleans students math and ELA test scores (not proficiency rates) in grades 4-8 improved by only 2.5 NCEs between 2008 and 2011 (similar to percentiles, NCEs range from 1 to 100 with a mean of 50, which for these data represents the Louisiana average score). This study only had data for New Orleans schools, so it could not produce value-added estimates for New Orleans schools compared with schools in the rest of the state.

Math and ELA Test Scores in Grades 4-8 (Not Value Added)

<table>
<thead>
<tr>
<th>NCE (50.0 = Louisiana Average)</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>59.8</td>
<td>61.3</td>
<td>63.9</td>
<td>64.6</td>
<td>61.4</td>
<td>60.8</td>
<td>63.8</td>
<td>65.3</td>
</tr>
<tr>
<td>ELA</td>
<td>50.9</td>
<td>55.0</td>
<td>53.7</td>
<td>52.3</td>
<td>54.0</td>
<td>57.3</td>
<td>58.3</td>
<td>56.2</td>
</tr>
<tr>
<td>OPSB Charter</td>
<td>59.8</td>
<td>61.3</td>
<td>63.9</td>
<td>64.6</td>
<td>61.4</td>
<td>60.8</td>
<td>63.8</td>
<td>65.3</td>
</tr>
<tr>
<td>OPSB Direct</td>
<td>50.9</td>
<td>55.0</td>
<td>53.7</td>
<td>52.3</td>
<td>54.0</td>
<td>57.3</td>
<td>58.3</td>
<td>56.2</td>
</tr>
<tr>
<td>New Orleans</td>
<td>44.2</td>
<td>43.7</td>
<td>45.1</td>
<td>46.8</td>
<td>44.4</td>
<td>43.3</td>
<td>44.4</td>
<td>46.9</td>
</tr>
<tr>
<td>RSD Charter</td>
<td>38.7</td>
<td>41.7</td>
<td>43.8</td>
<td>44.3</td>
<td>39.3</td>
<td>41.5</td>
<td>43.0</td>
<td>43.8</td>
</tr>
<tr>
<td>RSD Direct</td>
<td>31.9</td>
<td>31.7</td>
<td>33.3</td>
<td>34.6</td>
<td>28.3</td>
<td>30.6</td>
<td>31.7</td>
<td>34.5</td>
</tr>
</tbody>
</table>

Source: McEachin et al., Tables 2 to 5. All students in the sample had test scores from the previous year. Student test scores (iLeap, Leap and GEE exams) are standardized by grade, subject and year using the statewide student-level means and standard deviations, but they are not value-added estimates of school effectiveness. Standard deviations converted to NCEs by the AFT.
• The achievement changes in the USC study are half the size of the impact of class-size reduction found in the Tennessee STAR experiment, and in the same range of achievement gain as comprehensive school reforms such as Success for All. This raises the question of why the New Orleans school system was torn apart by a costly and unproven experiment when there already were equally effective and proven reforms such as class-size reduction and Success for All that could have been implemented quickly with far less disruption and at a fraction of the cost.

• **Growth in passing rates does not measure school effectiveness.** Trends in school proficiency rates are used to support the claim of a New Orleans miracle. As predicted by the research literature, one high-quality study found that some New Orleans schools with big proficiency gains were no more effective—based on value added metrics—than schools with low growth in proficiency rates. Some schools with little proficiency growth had high value-added scores. Similarly, CREDO’s sophisticated study of school closures in New Orleans since 2010 found that before closure, “learning gains” for students in the closed schools were actually similar to comparable RSD students even though labeled chronically failing by criteria that included changes in proficiency metrics. Proficiency trends are more likely to measure demographic change and other differences between successive cohorts of students cycling through a school than the performance of the school.

• **Recent school closings resulted in negative impacts on achievement.** New Orleans has become the mode for portfolio school districts. A key component of the portfolio district concept is the closing of “failing” schools and replacing them with a charter school, or switching failing charter schools to another operator. In 2010, New Orleans received a $500 million federal grant to expand its program of closing low-scoring schools. Under the terms of the Investing in Innovation (i3) validation grant, Stanford researchers performed an evaluation of the educational effectiveness of the school closures and replacements since 2010. Using quasi-experimental research techniques, the researchers found negative impacts on student achievement—usually at statistically significant and meaningful levels—for first cohort of schools. Students from closed schools suffered learning loss when they subsequently moved to other schools or were accepted for re-enrollment in the new school that was taking over the building.

• **Louisiana restricted data access to researchers it selected.** High-quality independent research requires access to high-quality data. Research on Reforms Inc. sued the Louisiana Department of Education to obtain the raw data used by a state-approved research organization to evaluate the New Orleans takeover. Initially, a state district court held that raw data were not a public record and LDOE had the discretion to select its evaluators. However, an appeals court unanimously reversed the district court ruling, declaring that decoded student data records are public records not protected by privacy laws under the Louisiana Public Records Act, and that the act does not allow a state agency to inquire as to the reason for the request of the public record.
ENDNOTES

1 Andrew Ho. “The Problem With ‘Proficiency’: Limitations of Statistics and Policy Under No Child Left Behind.” Educational Researcher, 2008, 37(6), 351-360. Ho demonstrates that the proficiency metrics offer only limited and unrepresentative depictions of test score trends, gaps, and gap trends. The limitations are unpredictable, dramatic and difficult to correct, and interpretation of proficiency metrics generally leads to incorrect or incomplete inferences about change.


4 Sacerdote, 2012.


6 McEachin et al., 2013.


8 McEachin et al., 2013. Test scores cited are converted to NCEs by the AFT, and they are not value-added scores.


12 CREDO, New Schools for New Orleans Year 2 Report, Hoover Institution, Stanford University, October 2013.


14 The portfolio district concept merges decentralization, charter school expansion, test-driven accountability, and the closing of “failing” schools. Low-scoring schools are either shut down, converted to charters, or for some charter schools, their charters given to another charter school operator promising better results.

15 CREDO, 2013.


17 www.researchonreforms.org/html/documents/AppealCourtReversesDistrictCourt.pdf