Redesigning the Annual NYC School Survey: Lessons from a Research-Practice Partnership

Lisa Merrill
Camille Lafayette
Shifra Goldenberg

February 2018
Redesigning the Annual NYC School Survey
Lessons from a Research-Practice Partnership

Lisa Merrill
Camille Lafayette
Shifra Goldenberg

February 2018
Redesigning the Annual NYC School Survey: Lessons from a Research-Practice Partnership

Schools are more than collections of individual students, parents, teachers, administrators, and staff. They are complex organizations, with webs of relationships, routines, structures, and environments (Cook & Yanow, 1996). An emerging body of research is showing that organizational characteristics—such as student-centered learning environments, instructional leadership, and safety—can influence teaching and learning, and might be an important lever for improving school quality and outcomes (Bryk et al., 2010, Kraft, et al., 2016). More and more school districts around the country, including Chicago, Philadelphia, and Seattle, are using surveys to gather information about these key organizational characteristics.

The NYC Department of Education (NYC DOE) was an early adopter of this strategy. Each year since 2006, the NYC DOE has distributed surveys to all students in grades 6-12, as well as all teachers and parents in the district. The NYC School Survey is administered in over 1,800 schools to nearly a million individuals each year, making it the largest annual education census in the United States.

Since 2010, the Research Alliance for New York City Schools has worked with the district to analyze and suggest improvements to the Survey. Our early work culminated in a report assessing the reliability and validity of the Survey’s measures and making a number of recommendations about how the Survey could be improved (Nathanson, et al., 2013).

In 2014, new leadership at the NYC DOE invited the Research Alliance to play a larger role in a comprehensive redesign of the School Survey. The redesign was an integral part of the DOE’s effort to articulate a clear and compelling framework for school improvement and support across the City. This eventually crystallized into the “Framework for Great Schools,” which outlines key elements of school climate and capacity that research has shown to be important for improving student outcomes. The Survey redesign was aimed at measuring the Framework’s central elements in

Nearly a million individuals in over 1,800 schools take the annual NYC School Survey.
each of the City’s schools—and then making this information available to district and school leaders, teachers, parents, and the general public.

The process of redesigning the NYC School Survey has been highly collaborative. The work has exhibited the shared learning and ongoing exchange of knowledge that is characteristic of a strong research-practice partnership. Throughout the multi-year (and still ongoing) effort, researchers and practitioners have worked together to identify and build useful measures of important school capacities. Researchers have had to be responsive to the district’s priorities and timelines to build a useful instrument. In turn, district leaders have given heed to the researchers’ guidance about building a coherent set of instruments that are grounded in sound theory and prior evidence. The result has been a stronger set of measures to identify schools’ strengths and weaknesses. Just as important, the partnership has established a process for continuing to assess the quality of the measures and modify them as needed over time.

While the redesign effort has been successful in many ways, it has also encountered challenges, especially in balancing the theoretical and methodological ambitions of the researchers against the practical and operational realities of a large, complex and heterogeneous school system. This document draws on our experience to help other school districts anticipate some of these challenges. We share lessons learned to date and highlight questions that remain about developing school surveys and using their results to inform research, policy and practice.

For further details about the process of redesigning the NYC School Survey, the technical properties of the new Survey, and a discussion of measurement strategies, please see *Redesigning the Annual NYC School Survey: Improving Measures of School Climate through a Strong Research-Practice Partnership.*

**Lessons for Survey Development**

**Timing counts: Bring relevant research to the table when policy decisions are being made.**

The original NYC School Survey was developed in 2005 and first administered in the spring of 2006. The instruments (i.e., surveys for teachers, parents and students) were based largely on surveys that had already been in use in the Chicago Public Schools for several years, as well as input from a small panel of district administrators
and researchers. As early as 2008, researchers identified significant limitations in the NYC Survey’s ability to capture valid and reliable information about school learning environments (Rockoff, 2008; Nathanson et al., 2013). Despite these assessments, the NYC DOE made only minor changes to the Survey, in the interest of maintaining continuity and the ability to compare schools on the same measures from year to year.

In 2014 a critical policy window opened, making a larger overhaul of the NYC School Survey possible. Newly elected Mayor Bill de Blasio had appointed Carmen Fariña, a long-time teacher and school administrator, as the Chancellor of New York City Schools. The incoming administration brought in a new set of education policies and priorities, which placed greater emphasis on assessing and nurturing schools' capacities to support good teaching and learning. The new DOE leadership team turned to prior research about the aspects of a school’s learning environment that were most important for producing better outcomes. They sought to redesign the Survey with a focus on these key aspects of school capacity. Because of the Research Alliance’s prior work on the Survey and familiarity with the relevant research, we were well positioned to assist.

While political shifts facilitated the Survey redesign in NYC, tackling the project at the beginning of a new administration also created its own set of challenges. In some cases, for example, we were developing survey items to measure aspects of a district strategy that were still being defined. This timetable had implications throughout the Survey redesign process, some of which we highlight below.

**Set the stage for collaboration between practitioners and researchers.**

The Survey redesign was led by a working group consisting of leaders from multiple NYC DOE departments and researchers from the Research Alliance. The DOE leaders brought deep expertise about the operations of—and policies driving—NYC schools. The researchers provided an evidence-based perspective on school improvement and school climate and capacity, as well as technical expertise related to Survey design.

At any given point, there were around 10 people involved with the working group. Over time, the group’s composition shifted slightly. For example, early on, as we were thinking about the Survey’s high-level goals, the DOE’s strategy team played a central role. Later, the working group shared drafts of the Survey across multiple offices and departments, and when we started thinking about survey analysis, the
DOE team focused on school performance became more actively involved. Having a small group of the “right” representatives—that is, people who were informed about relevant operations and policies and who had access to leaders with decision-making authority—was key to the project’s success.

Early working group meetings established ground rules to guide the collaboration. These included, for instance, a commitment to regular communication and adhering to strict timelines. We also agreed that the Survey was a work in progress and should be revised as needed over the next two or three years.

As our work continued, a healthy but challenging tension emerged between NYC DOE and Research Alliance working group members around how to balance the practical realities of survey implementation and district priorities with the desire to develop a technically rigorous instrument. Placing researchers and practitioners at the core of the redesign process allowed both of these perspectives to be voiced, and created opportunities for working group members to teach and learn from one another. Candid discussions helped build trust and understanding, and created a sense of mutual respect and accountability that served the group well throughout the project.

**Establish a shared vision for the survey’s purpose.**

The redesign effort reflected an important policy shift at the NYC DOE: Moving forward, there would be less emphasis on using School Survey information for accountability purposes (though this was not taken off the table entirely) and more emphasis on providing actionable feedback and targeted support to schools. The working group agreed that the new Survey should focus on collecting information that district and school leaders could use to assess the strengths and weaknesses of particular schools and to guide decisions about allocating resources and support.

Establishing a shared purpose for the redesign— and centralizing the project in a dedicated working group—helped address one notable challenge that had developed in earlier work on the Survey. Previously, many different DOE divisions had contributed ideas for what the Survey should measure. This produced a litany of disparate questions with limited cohesiveness and utility. Clarifying the Survey’s goals helped set important parameters for what to include—and not include—in the final product.
**Start with big-picture concepts.**

For the first several weeks, the working group focused on identifying the broad dimensions of school climate that should be captured on the Survey. We did this by asking: (1) What do district and building leaders need to know about the climate and capacities of their schools? And (2) of these aspects of school climate and capacity, what can be measured most effectively through a survey?

For example, resource management is certainly an important element of school capacity. However, this can be assessed by reviewing budget documents and allocations of staff time. On the other hand, the degree of trust between a principal and teachers (or between students and teachers) is an important dimension of school capacity that most likely can only be measured by asking members of the school community about their experiences and perspectives.

In determining which aspects of school capacity to include on the Survey, the working group relied heavily on school improvement research that had been conducted by the University of Chicago Consortium on School Research (Bryk, et. al., 2010). We also drew on an earlier memo the Research Alliance had written to the NYC DOE outlining a variety of suggestions for revising the original Survey. In addition, the researchers undertook a careful review of other literature about efforts to assess school climate and functioning. From all of these sources, we compiled a list of key dimensions of school capacity that the NYC School Survey should measure. It’s important to note that, at this point in the process, we focused on broad concepts and avoided thinking about individual survey items. This kept us from getting sidetracked by pinpointing or rewording specific items. The measures we identified as important were ultimately codified in the NYC DOE’s Framework for Great Schools (see Figure 1 on page 6).

**Build from existing measures and survey items.**

Once we had identified the basic dimensions we wanted to measure, we took advantage of the fact that many school districts—including the NYC DOE—had already used surveys to assess school climate. Before writing any new items, we conducted a thorough scan of available survey measures (i.e., collections of individual survey items grouped together to measure an underlying concept). Our sources included previous iterations of the NYC School Survey; Chicago’s surveys; national surveys; and other, prior research studies. In total, we were able to identify 29 previously validated measures (i.e., measures that have been used on surveys of
REDESIGNING THE ANNUAL NYC SCHOOL SURVEY: LESSONS LEARNED

similar populations and found to be reliable measurement tools), which we decided to include in the redesigned Survey. Previously validated measures have a number of advantages. They have already been tested and refined to ensure they meet basic standards of good survey measurement (discussed further on page 12). Furthermore,

Figure 1: The Framework for Great Schools

The 6 elements are comprised of 18 indicators each with specific measures which, together, predict growth or stagnation in schools

<table>
<thead>
<tr>
<th>Element</th>
<th>Indicator</th>
<th>What good looks like</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigorous Instruction</td>
<td>Course clarity</td>
<td>High standards are set in every classroom; with a focus on customized, inclusive and motivating instruction, as well as active student engagement developing critical thinking skills</td>
</tr>
<tr>
<td></td>
<td>Quality of student discussion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>English and Math instruction</td>
<td></td>
</tr>
<tr>
<td>Supportive Environment</td>
<td>Safety and order</td>
<td>A school culture where students feel safe, supported and challenged by their teachers and peers, and are engaged in ambitious intellectual activity</td>
</tr>
<tr>
<td></td>
<td>Social emotional learning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Academic support and press</td>
<td></td>
</tr>
<tr>
<td>Collaborative Teachers</td>
<td>Culturally responsive pedagogy</td>
<td>Teachers committed to success and driven to improve in their classroom and across the school; school leadership pays deliberate attention to professional growth; culture of continuous improvement and developing a school-based professional community</td>
</tr>
<tr>
<td></td>
<td>Professional development</td>
<td></td>
</tr>
<tr>
<td></td>
<td>School commitment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Professional community</td>
<td></td>
</tr>
<tr>
<td>Effective School Leadership</td>
<td>Inclusive/facilitative leadership</td>
<td>Principals leading change at the community level, nurturing the leadership development of others, and providing guidance over time to sustain a coherent instructional program</td>
</tr>
<tr>
<td></td>
<td>Instructional leadership</td>
<td></td>
</tr>
<tr>
<td>Strong Family-Community Ties</td>
<td>Parent involvement</td>
<td>School leadership drawing on the resources within the building and from the local community, encouraging partnerships with families, local business, community organizations, and city agencies</td>
</tr>
<tr>
<td></td>
<td>School-community partnerships</td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td>Family-staff trust</td>
<td>Across all relationships, there is respect, personal regard, assumed competence, and integrity; and all parties value and respect each other</td>
</tr>
<tr>
<td></td>
<td>Central trust</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student-teacher trust</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Staff trust</td>
<td></td>
</tr>
</tbody>
</table>

Source: The NYC DOE.

Notes: The 2010 book *Organizing Schools for Improvement: Lessons from Chicago* (Bryk et al., 2010), developed by the University of Chicago Consortium on School Research, provided an important basis for NYC's survey redesign effort.

Bryk's book elucidated a set of factors that had been linked to improvement in Chicago Public Schools. It also offered survey tools for measuring those factors. The NYC DOE aimed to build on this framework and modify it for NYC's unique context, including updating the measures from Bryk's work (which were from the 1990s and early 2000s). The district also sought to draw on other available research about the elements of school climate and capacity that are important for improving student outcomes.

The result of these efforts was the creation of NYC's Framework for Great Schools. Figure 1, above, shows the six elements that make up NYC's Framework, along with the survey indicators designed to capture information about each of these elements. Measures are nested within indicators. The Framework has 32 measures that are nested in the 18 indicators shown above. To see the full list of measures, see Appendix C.

Learn more about the NYC Framework for Great Schools here:
http://schools.nyc.gov/AboutUs/schools/framework/default.htm
these measures are often being used in other contexts, which creates the ability to compare survey results with data from other places.

For some aspects of school capacity that we wanted to capture, there were no previously validated survey measures—not surprising given our goal of tailoring the Survey to NYC’s specific, current context. In these cases, we created new items, and committed to careful ongoing analyses of their reliability and validity once the data were collected. Following well established guidelines for constructing survey items, we aimed to write clearly and concisely, using language that would be familiar to the target audience. Ultimately, we produced three new Survey measures.

Figure 1 on page 6 provides an overview of the NYC DOE’s “Framework for Great Schools” and a list of the survey measures—referred to as “indicators” in the figure—that were used to assess each element of the Framework.

Collect feedback from groups who will take the survey and use its results.

While the working group represented both operational expertise about NYC schools and technical skills in survey design, we recognized the importance of gathering feedback from other key stakeholders. As noted above, the NYC DOE shared early drafts of the Survey internally. This included district leaders and offices overseeing English language learners, special education, instruction and professional development, post-secondary readiness, guidance, parent involvement, and social emotional learning, as well as the district’s Research and Policy Support Group.

The Survey working group also collected feedback from teachers, students, and parents through a series of focus groups. To recruit focus group participants, the Research Alliance partnered with the United Federation of Teachers (UFT), the NYC DOE (particularly the Division of Family and Community Engagement, which connected us with parents), and Coro’s Youth Leadership Academy (which helped us engage a group of high school student leaders). We structured the focus groups to
elicit feedback about their overall experience with the Survey, the clarity of specific items, and whether the response scales being used were intuitive.

Receiving feedback on previously validated survey measures introduced yet another tension into our work. While edits could make the Survey more specific to the New York City context and clearer for respondents, making changes to previously validated items might affect the quality of those measures. This reinforced the importance of ongoing analyses to validate the measures, with a special focus on those that included new or modified items.

Lessons for Evaluating the Quality of Survey Data

Assess response rates and their implications.

The redesigned Survey was administered for the first time between December 2014 and March 2015. This initial administration resulted in completed surveys by 66,108 teachers (81 percent of all teachers); 459,929 parents (50 percent of all parents); and 425,387 students (81 percent of all students in grades 6-12) across the NYC public school system.

Several factors should be taken into account when assessing the adequacy of survey response rates. The first is the absolute percent of the population that actually completed the Survey. In general, the higher the response rate, the more likely it is that the results fully represent the perspectives of the larger population. Researchers commonly consider a response rate of 70 percent or more to be an indicator that a survey captures the views of a broad sample of the population of interest. However, another important factor is whether those who responded to the Survey are likely to have systemically different views than those who did not.

The fundamental goals of a survey also matter when we consider response rates. When the goal is to examine average school conditions across a system, then system-wide response rates may be most meaningful. But, if the goal is to learn about the relative strengths and weakness of individual schools, then each school’s response rate is important. Because the NYC School Survey was designed to gather information about both the overall system and individual schools, response rates at each level matter, as does the variation of response rates across schools (i.e., do some kinds of schools have higher response rates than others?). (See Part II of Redesigning the Annual
NYC School Survey: Improving Measures of School Climate through a Strong Research-Practice Partnership for a breakdown of NYC response rates at each of these levels.)

Particularly when response rates are low, it is important to ask whether the people who took the Survey were meaningfully different from those who didn’t. For example, in NYC, we found that, in schools with low response rates, students who took the Survey and those who did not were similar in many ways—but, perhaps not surprisingly, students who did not take the Survey were more likely to be chronically absent. This does not mean that our results are necessarily skewed, but it is an important reminder to consider the voices that may be missing from our survey results.

Several statistical methods are available to make adjustments to survey results, based on potential bias from low or differential response rates. It is crucial, however, to

**Survey Administration**

Successful administration of a district-wide survey requires financial resources and careful planning. Questions to consider include:

- How will the district promote and publicize the survey to encourage as many people as possible to respond?
- In what languages should the survey be offered, and how will it be translated?
- Should the survey be administered on paper, digitally, or both?
- Will the survey be fully anonymous or include an identifier that can link results to administrative data for analysis?
- How long should the survey be “in the field” (i.e., how much time should people have to respond)?
- Who is responsible for the day-to-day tasks of distributing the survey, responding to questions about it, and collecting it? How will they be trained?

The NYC DOE has been able to dedicate significant financial and human resources to administering the School Survey. The DOE contracted with a consulting firm, KPMG, to print, distribute, and score surveys. The district also trained a staff member at every school about the logistics and ethics of survey administration. Each school identified a parent survey coordinator to help boost response rates.

School leaders were given the option of having students complete their survey on paper or online. In either case, students completed their survey during school hours. For parents, KPMG mailed a paper survey to each student’s household; parents could also choose to take their survey online. The parent and student versions of the School Survey were translated into a variety languages, including Arabic, Bengali, Chinese, French, Haitian Creole, Korean, Russian, Spanish and Urdu. The teacher survey was administered online at the teachers’ convenience at home or in school.

Without these investments, it is unlikely that the NYC DOE would have achieved the response rates that it did. Districts should be conscious of the complexity and likely costs of survey administration.
build the strongest possible infrastructure for obtaining high response rates from the people whose perspectives are most important to the learning agenda.

**Assess the reliability and validity of the survey measures.**

One of the early commitments made by our working group was that the 2014-2015 Survey was a work in progress, and that we would carefully assess the reliability and validity of the measures after survey results were available. In keeping with this goal, the Research Alliance conducted a number of statistical tests to determine the quality of our survey measures (see textbox on “Survey Measurement Properties” below), including those that had been validated in previous research. We shared the results of these tests with the NYC DOE as quickly as possible, to inform potential changes to the next year’s Survey. (See Part III of *Redesigning the Annual NYC School Survey: Improving Measures of School Climate through a Strong Research-Practice Partnership* for details.)

In general, we found that the redesigned Survey did a much better job of capturing distinct aspects of school climate and capacity than previous iterations. Many measures used in 2014-2015 were of very high quality, but there was room for improvement in others. Using the results of these analyses, the working group made several revisions to the 2015-2016 Survey. We also continued holding focus groups with key stakeholders to get their feedback on both the original items and those that were added or revised. Based on preliminary analysis of the 2015-2016 Survey, these changes further improved the quality of the Survey, giving us greater confidence in the information the Survey was gathering about school climate and capacity.

For the measures that were changed, however, we will not be able to directly compare results across years. As with the original redesign, the difficult tradeoff between efforts to improve measurement and the value of continuity is a significant challenge. Explicit conversations can help researchers and practitioners make good decisions about how to handle this tradeoff.
Lessons for Survey Analysis

Select an analytic strategy that matches the intended use of your survey results.

In New York City, we have explored two main strategies for analyzing survey results, each of which serves a different purpose.

- **School-Level Averages.** Survey results from each school (either within or across respondent types) can be aggregated into a school-wide average “survey score” for each measure. This strategy, which is employed frequently by school districts, can be used to rank and sort schools. The information can help districts identify schools in need of support, and can let the public know about where a particular school stands within a broader spectrum. For researchers, this strategy allows for analysis that examines the relationship between school-level measures of climate and capacity with other school-level outcomes (e.g., attendance rates or test scores).

- **Variation Within Schools.** One limitation of the approach described above is that not all students, teachers, and parents have the “average” experience in a school. Describing the range of experiences in a school can help school leaders identify particular groups who might benefit from specific supports or interventions (Langley et al., 2009). Variation can be described by creating individual-level averages for each measure, and then graphing a distribution of all responses. Another strategy is to create “profiles” of survey respondents who responded similarly to a survey. This can help school leaders think about why different respondents are falling into these categories, and experiencing the school context differently.

The NYC DOE has focused primarily on school-level averages and percentages in reporting and using data from the NYC School Survey. This is a sensible strategy for identifying schools within the system that may be in need of improvement and support. These school-level averages and percentages also provide a transparent and accessible summary of student, teacher and parent perceptions of the learning environment at each school. That said, the Research Alliance is committed to helping the district build more nuanced indicators, including measures of within-school variation. Adding these kinds of measures to reports may help schools better understand their internal dynamics and better meet the needs of their many diverse stakeholders.
Concluding Thoughts

In light of the Every Student Succeeds Act’s requirement that states collect and report data on at least one “alternative” measure of school quality—and as evidence mounts that certain organizational characteristics are important for student outcomes—it is likely that school survey efforts around the country will continue to generate momentum.

We hope that this document illustrates how close collaborations between researchers and school districts can help create higher-quality survey instruments. We also hope we have highlighted some of the key challenges to expect and the questions to consider when approaching this work.

For further information about all of the topics discussed here, and for a copy of the 2015-2016 NYC parent, student, and teacher versions of the School Survey, please visit Redesigning the Annual NYC School Survey: Improving Measures of School Climate through a Strong Research-Practice Partnership.

Survey Measurement Properties

After the first administration of the new Survey, we examined the following aspects of measurement quality:

- **Reliability**: Do individuals tend to respond to survey items within a measure similarly? (This tells us whether the questions in a measure are looking at a coherent concept.)
- **Within-School Agreement**: Do individuals within a school tend to reply similarly to items in a given measure? (This helps us learn whether the measure captures something about the school, rather than about the individual.)
- **Responsiveness**: Are responses to a measure clustered toward the very top or bottom of a response scale? (This tells us whether the measure will be able to capture changes over time, or if there’s little room for higher/lower responses. It may also reveal if the response scale has insufficient nuance.)
- **Face validity**: Does the measure “look like” it’s going to measure what it’s supposed to, to the eyes of relevant stakeholders?
- **Content validity**: Does the measure seem to include appropriate items that capture all relevant elements of the measure being measured?
- **Concurrent validity**: Does a measure correlate with another, already established measure of the same or a closely related concept, administered at the same time?
- **Predictive validity**: Does a measure forecast a future event, behavior, or outcome?

For more details on the technical properties of the Survey, and the statistical methods used for this analysis, please see Part II of Redesigning the Annual NYC School Survey.

For more information on types of validity, see: Weiss, 1998 (p.144-146).
**Endnotes**

1 Broadly speaking, reliability is the extent to which a survey instrument produces consistent results each time it is used. Validity is the extent to which an instrument measures what it is supposed to measure. For more information, see the “Survey Measurement Properties” textbox on page 12.


3 The new measures we developed were about Social-Emotional Skills, Guidance, and School Leadership Team (a specific committee that every NYC school is required to have). Please see the full technical document for more details.

4 Due to timing constraints for the parent and student surveys, which needed to be translated and printed, these focus groups took place after 2014-2015 survey administration, and feedback was incorporated into the following year’s survey.

5 Because the NYC School Survey seeks the perspectives of parents, students, and teachers, this 70-percent threshold applies to each of three groups.

6 Meaning, other than academic outcomes.
The Research Alliance for New York City Schools conducts rigorous studies on topics that matter to the city's public schools. We strive to advance equity and excellence in education by providing nonpartisan evidence about policies and practices that promote students' development and academic success.