Redesigning the Annual NYC School Survey
Improving Measures of School Climate through a Strong Research-Practice Partnership

*Part IV: Exploring Two Approaches to Analyzing Survey Results*

Lisa Merrill
Camille Lafayette

February 2018
Contents

Redesigning the Annual NYC School Survey: Part IV: Exploring
Two Approaches to Analyzing Survey Results: School-Level Scoring
and Individualized Scoring Within Schools .......................................................... 1

Strategy I: School-Level Estimates ........................................................................ 2

Strategy II: Within-School Variation ........................................................................ 6

Discussion .................................................................................................................. 10

Endnotes .................................................................................................................... 11

References .................................................................................................................. 12
Redesigning the Annual NYC School Survey: Part IV

Exploring Two Approaches to Analyzing Survey Results: School-Level Scoring and Individualized Scoring Within Schools

The strategy we use to analyze survey results has implications for how we can use the information. In this section of “Redesigning the Annual NYC School Survey,” we describe two different techniques that can be used to analyze school survey results, and discuss the benefits, drawbacks, and implications related to each.

One of the most common survey analysis strategies is to aggregate individual-level surveys to the school level to create a single number that represents how the school scores on each measure or element. Aggregating to a single score is common because it easily ranks and sorts schools, by measure or element, into higher and lower scoring schools. In turn, districts can use these statistics to determine which schools are stronger and which are weaker on different measures and elements, which can inform decisions about where to allocate resources and support. Schools can also use the results to see how they are doing compared to other schools. Being able to identify targeted areas of improvement may help schools invest more resources and supports where needed. Lastly, this strategy enables researchers to explore the relationships between school-level climate measures and other school-level outcomes, like achievement or attendance.

A significant downside to aggregating a school’s climate to a single number is that not all students, teachers, and parents experience the school climate the same way. Organizational theory suggests that people within organizations often view the organization differently, depending on their roles, responsibilities, or identities (Senge, 1997; Deming, 1990). People often form subgroups, which leads to the existence of “microclimates” within an organization (Rowan et al., 1991). Microclimates can be created by formal structures: for example, the English department may have a different view of leadership than the Math department, or the sixth-grade team may have a different perspective than the eighth-grade team. Microclimates may also develop less formally, for example, among a group of
teachers who have similar pedagogical perspectives or a group of new teachers who 
socialize together. Describing this range of experience could be enlightening to 
practitioners and helpful for school improvement.

This suggests the need for another strategy to analyze survey responses—one that 
focuses on the variation of individual responses within a school. Such variation can 
help us assess whether there are subgroups (i.e., those who might respond particularly 
positively or negatively to a specific measure) who might benefit from specific 
supports or interventions to improve their experience in school. This analytic strategy 
would provide principals with information about high and low scoring survey 
responses within their own schools, which could allow school to hypothesize why 
some students, teachers, or parents feel more positively or negatively about the 
school climate and then test a solution to improve that group’s experience (Langley 
et al., 2009).

Below, we walk through these two strategies for analyzing survey results, using the 
Student-Teacher Trust measure from middle schools as an example. The goal is to 
demonstrate the differences between the strategies and implications for using either 
strategy. However, both approaches could be applied to any of the survey measures. 
Using the strategies together could provide valuable information about how a school’s 
average scores compare with others in the system, as well as the (sometimes wide) 
variety of perspectives within a given school.

**Strategy I: School-Level Estimates**

The first strategy is to calculate a single school-level estimate that represents the 
climate for the entire school. There are two popular ways to aggregate scores: an 
average score\(^1\), or the average percent positive. For example:

- An average score would indicate that in School A, on average, students feel 
  that Student-Teacher Trust is a 3.1 (on a scale from 1 to 4).

- A percent positive would indicate that, at School A, 80 percent of students 
  responded positively\(^2\) to the items in the Student-Teacher Trust measure.\(^3\)

While these two strategies are similar because they result in a single estimate that 
represents how most students feel, percent positive can be more intuitive to 
education stakeholders. The NYC DOE uses the percent positive to report results of
the Survey on district School Quality Reports. Stakeholders can use these school-level scores to compare different schools and to identify areas where individual schools should focus improvement efforts.

School-level scores can also be used to learn about conditions across the district. Table 2 below illustrates this by providing some descriptive statistics for the Student-Teacher Trust measure. As noted above, this measure uses a 1-4 scale, where 1 = “Strongly disagree”, 2 = “Disagree”, 3 = “Agree”, and 4 = “Strongly agree”. It is important to understand that we are talking about two types of averages in this section. First, the school-wide average for a measure is the average of all individual responses on a specific measure within a given school. Second, the district-wide average is the mean of all the school-wide averages across the district. The average shown in Table 2 (3.14) is the district-wide average for all middle schools.

The table provides three different ways of looking at the variation around this district-wide average. The “Std” column shows the standard deviation (.18), which is based on the distance between each school and the district wide average. The smaller the standard deviation, the more similar schools are to the average. In the two right-hand columns, we provide the minimum and maximum school-level averages from across the district. In the two middle columns, we show the score that is equivalent to the 10th percentile and the 90th percentile, which eliminates the most extreme values. For each pair of numbers, if the spread between the two is narrow, we can assume that there were not large differences in school-level averages across the district. Conversely, if the spreads are wide, we can assume that the differences in school-level averages across the district are large.

### Table 2: Descriptive Statistics from Middle School Student Survey 2014-2015

<table>
<thead>
<tr>
<th></th>
<th>District Average</th>
<th>Std</th>
<th>P10</th>
<th>P90</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student-Teacher Trust</td>
<td>3.14</td>
<td>0.18</td>
<td>2.95</td>
<td>3.40</td>
<td>2.64</td>
<td>3.83</td>
</tr>
<tr>
<td>Sample size</td>
<td>281</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Source(s):** Research Alliance calculations based on data obtained from the NYC Department of Education.

**Note:** School-wide average scores are created using individual responses by first averaging each student’s responses to the items within a measure and then averaging all students in a school.
Determining if a spread is narrow or wide is not an exact science. When assessing the spread, we suggest readers use the scale to determine if the spread is meaningful. When a spread spans multiple response categories, it is generally considered wide. For example, if an average moved from near the “Strongly disagree” response category to the “Agree” category, that spans three response categories and would be considered a wide spread. On the other hand, if the average moved slightly within the “Strongly disagree” response categories, the spread would be considered narrow.

The reason why it is important for readers to understand the spread of measures is two-fold. First, readers should think critically about statistics and what they may mean practically, at the ground level. For example, imagine a very narrow spread, say from 3.0 and to 3.1 on a 1-4 scale. Would we expect the school scoring the minimum to look and feel very different from the school scoring the maximum? For most people, the answer would be no, meaning this is not a particularly meaningful difference. The same logic may be applied to a spread that is wide. For example, if we saw a spread from 1.5 to 4.0, most people would expect to see a tangible difference between the minimum school versus the maximum school. The second reason why spread matters is that it can tell us something about the quality of the measure. In some cases, it is possible that a measure is not fine-tuned enough to pick up observable differences between schools. For example, we might know that there are observable differences in how safe schools are, but find that our safety measure is unable to pick up on such differences—resulting in a narrow spread. If the district-wide differences on a measure are not meaningful, then we may want to consider using alternative measures to compare schools.

The Importance of Precision

Another issue to consider when creating school-level estimates is the amount of precision and/or error when calculating scores. For example, in schools with fewer respondents, the school-wide score is more susceptible to change based on one or two respondents’ perspectives. Researchers think of these schools as having less precise estimates. Schools with larger number of respondents have much more precise estimates. More advanced statistical methods can rescore school-level estimates based on each school’s precision, but the use of these methods can make the scoring more abstract and confusing for stakeholders. We hope to explore what this scoring may look like and how it would be interpreted by practitioners in future work.
In general:

- When the spread is narrow, it often makes sense to focus on the tails, because most people in the middle have almost the same scores, and the only meaningful differences are at the extreme values.

- When the spread is wide, there may be meaningful differences across the entire sample.

Looking back at Table 2, we see that the district-wide average for Student-Teacher Trust is 3.1, with the lowest school average is 2.7 and the highest school average is 3.8. This spread is pretty wide, spanning several categories on our 1 to 4 scale (2.7 sits between “Disagree” and “Agree”, while 3.8 is near “Strongly agree”). When we compare the 10th and 90th percentile schools, we eliminate any real outliers. In this case, the 10th percentile school is 2.9 and the 90th percentile school is 3.4, a difference that is much smaller (.5), but still travels from below the “Agree” category to halfway between “Agree” and “Strongly agree.” In our view, this is still meaningful.

Another way to look at variation across the district is to plot all school-level averages in a graph. Figure 5 below shows each middle school’s average on the Student-Teacher Trust measure, sorted from lowest to highest. The figure reveals that many schools in the middle have relatively similar averages, but there are some significantly higher and lower averages at either end. Schools at the left side of the figure may need additional support to build Student-Teacher Trust, while schools on the right-hand side might be model schools that others could learn from. In short, this graph can help us identify schools with very strong and very weak Student-Teacher Trust, but it also suggests that there are not meaningful differences among the majority of schools in the middle.
Figure 5: School Averages for Student-Teacher Trust in NYC Middle Schools, 2014-2015

Source(s): Research Alliance calculations based on data obtained from the NYC Department of Education.

Note: Data is from 281 middle schools. School-wide average scores are created using individual responses by first averaging each student’s responses to the items within a measure and then averaging all students in a school.

Strategy II: Within-School Variation

While the school-level average is helpful for comparing schools to one another, it does not necessarily paint a complete picture of the climate in any given school. We know that not all individuals in a school have the “average” experience. In the case of Student-Teacher Trust, some students may feel like they have better relationships with teachers than indicated by their school’s average, and some may feel like they have worse relationships with their teachers than shown in their school’s average. Figure 6 below shows the range of individual student scores across all NYC middle schools. Each vertical line represents a school and the orange dots represent the school-wide average (the orange dots match Figure 5 above). The top (green) dot represents the 90th percentile response for the school. The bottom dot (blue) represents the 10th percentile response for the school. The schools are ordered identically to Figure 5. This type of within-school analysis requires individual-level data, which the Research Alliance has, but the NYC DOE does not. (See Part II of Redesigning the Annual NYC Survey for more information about this issue.)
Figure 6 shows that in most schools, there is quite a lot of variation in how students think about Student-Teacher Trust. We can take this insight further and try to identify common “profiles” that show how clusters of students responded to this measure. Using a statistical method called Latent Class Analysis, which groups students based on the patterns of their responses, we can get a sense of how different groups of students feel about Student-Teacher Trust.

We found four common profiles for Student-Teacher Trust in middle schools. Students in the very high group responded that they “Strongly Agree” with almost all of the items. Students in the high group responded that they “Agree” or “Strongly Agree” to the items. Students in the low group responded that they had a mixture of

Source: Research Alliance calculations based on data obtained from the NYC Department of Education.

Note: Data are from 281 middle schools. School-wide average scores are created using individual responses by first averaging each student’s responses to the items within a measure and then averaging all students in a school. Top decile is the 90th percentile within each school, and lowest decile is the 10th percentile within each school.
“Agree” and “Disagree” responses. Students in the lowest group responded “Disagree” or “Strongly Disagree” to all items.

Figure 7 below shows the proportion of students in each school who fall into each of the four profile categories. The schools in this figure are sorted in the same order as in Figure 5 and 6 (using orange dots to illustrate the school-level average). As in Figure 6, each vertical line represents a single middle school.

This graph illustrates several characteristics of the Student-Teacher Trust measure. First, all schools have some students in the Very High, High, and Low groups. Second, schools with lower averages have a higher proportion of students in the Very Low and Low groups. Lastly, schools with higher averages have a smaller proportion of students in the Very Low or Low group and a much higher proportion of students in the Very High group.

Analyzing the within-school variation suggests that students who attend the same school often have different perspectives of the school climate. This phenomenon is also true for teachers and, to a lesser extent, parents. Providing schools with information about how many students, parents, or teachers feel very positively, very negatively, or somewhere in between may help school leaders begin to hypothesize why people are having different experiences in school. The goal is not to identify individual students, but to assess what may be happening in the school that is creating distinct groups, and to inspire thinking about strategies that might be used to improve the microclimate for students who are having a less positive experience. This type of analysis can help schools identify areas for improvement as well as areas to celebrate. Almost all schools in Figure 7 had at least some students who reported very positive and trusting relationships with their teachers. The key is to try and make that a reality for all students.
Figure 7: School Averages for Student-Teacher Trust by Profile in NYC Middle Schools, 2014-2015

Source: Research Alliance calculations based on data obtained from the NYC Department of Education.

Notes: Data is from 281 middle schools. School-wide average scores are created using individual responses by first averaging each student's responses to the items within a measure and then averaging all students in a school. The schools are ordered from lowest to highest average school score on Student Teacher Trust.
Discussion

Practically, how we analyze and display data depends on our intended purpose. If our goal is to rank or compare schools, then aggregating data to the school level is likely the most effective method. If a district wants to use survey data to understand the diversity of experiences within schools, then analyzing variation within a school is best. Of course, a district could pursue both of these strategies. However, in New York City, because the NYC DOE does not receive individual-level data (as discussed in Part II, examining variation is more difficult. The district would need to pre-specify profiles and ask KPMG to provide an additional analysis.

From a research perspective, these two strategies spark many questions. For school averages, we want to know how well these averages predict future student achievement, and how well they distinguish between schools. Are there thresholds that we should be aware of (i.e., when schools reach a certain level on a particular measure or set of measures, they exhibit more success than schools that do not achieve that level)? For within-school variation, there are many areas ripe for future exploration. For example, we wonder which students are more likely to be in a low or very high group, and how being part of that group predicts other student outcomes. Understanding how practitioners interpret and use these different types of survey findings is an important next step for this work.

Many states and districts around the country are beginning to administer school climate surveys to teachers, students, and parents. It is important that the data derived from these surveys be analyzed and reported in a way that supports the surveys’ purposes and provides actionable insights for stakeholders. This is fertile terrain for researchers and practitioners to work together, bringing different kinds of expertise to bear on survey design, analysis and reporting.

Check out the rest of our Redesigning the Annual School Survey collection.
Endnotes

1 Others may use model-based school-level estimates, employing multi-level modeling and/or Item Response Theory/Rasch scoring.

2 A positive response on a 4-point item would be selecting a 3 or a 4, for example, indicating “Agree” or “Strongly Agree.”

3 Student-level responses is the average of the response each student gave to each individual item in a measure.

4 First, this analysis creates an average score for each student—i.e., the average of their responses to each item within a measure. Then the analysis calculates the average of the individual average scores to create a school-level average.
References


