

Measuring School Quality Beyond Test Scores

FINAL REPORT
June 30, 2015

Principal Investigator:

Jack Schneider
College of the Holy Cross
jschneid@holycross.edu

Co-Investigators:

Hunter Gehlbach
Harvard University

Rebecca Jacobsen
Michigan State University



Executive Summary

In September of 2014, our research team contracted with the Somerville Public Schools to build a new measure of school quality—one that would go beyond test scores to capture a fuller range of outcomes and better reflect the work being done inside the schools. By providing more useful data to school and district leadership, we hoped to facilitate the setting and tracking of clear, holistic, measurable goals. And by packaging the new data in a straightforward manner, we hoped to reach an even broader audience—engaging parents and community members, giving them more usable information, and empowering them to advocate for children in the city.

Rather than merely assembling available data and then working backwards to determine their significance, our team began by building a school quality framework. Drawing on national polling and the prior research of our team members, we built a draft framework that we further refined through surveys and focus groups with key stakeholders in the city—teachers, parents, administrators, and community members. Particular efforts were made in these early stages to reach out to non-dominant populations, including translation of surveys into Spanish in order to reach the largest non-English speaking demographic in the city.

Once we had determined what to measure, we then set to work on how to measure. Some of the data for our framework could be acquired merely by querying the district. But much of it could only be collected by asking teachers and students directly. In many cases we adapted existing survey scales that had been validated by prior research. And in other cases, we drew on best practices to develop new survey scales. Ultimately, we developed a set of computer-based surveys taken by all students and teachers in the district.

While our primary survey collection method was through district-organized, computer-based surveys delivered through SurveyMonkey, we also piloted a low-cost and high-fidelity method of survey administration via cellphones. A sizable minority (12.5%) of teachers participated in a yearlong survey delivered via a smartphone application. And a much smaller subset of students (roughly 80 total) participated in surveying via text message. Response rates for both groups—approximately 60-75%—were highly encouraging.

Having collected all of this new information about the schools, we worked to create a user-friendly interface for displaying it. Working with two web developers, we designed an easy-to-use web portal that houses the data for each school—a tool that can be used by the school committee, district leaders, teachers, parents, and other interested parties. The tool has a number of unique features, including the use of color-coded “zones” that help users interpret scores, clear and straightforward narratives, and a highly visual interface.

Particularly worth noting is the fact that this tool was created in order to support schools. The tool is not designed to rank or label schools, and its design reflects this. Schools cannot be sorted into hierarchical lists, nor are they branded by levels, as they are by the

Measuring School Quality Beyond Test Scores

Final Report

state. Instead, each school has its own page that offers a full range of data, along with guides about where schools need to grow in order to reach community-set targets. Thus, we believe we have not only surpassed the existing state data portal in the range of data collected and presented, but also in the degree to which data will encourage engagement rather than shopping or competition.

Finally, with financial support from the city's SomerPromise office, we conducted an end-of-project poll to empirically test the new data system—an experiment involving 45 demographically-representative community members.¹ Results from that poll indicate clearly that users of our new web tool are both more informed about the schools and more confident in the quality of the schools than are users relying on existing data provided by the state. This makes a powerful case for making the tool available to the public.

This project has accomplished all of its initial aims. And if the district and school committee do nothing other than continue to administer surveys and feed data into a public-facing website, we believe that this project will be an unqualified success.

But we also believe that this work can continue to bear fruit for the city and its schools, particularly if modest investments continue to be made. In light of that, we offer several recommendations for future efforts. More detailed recommendations are included in the report that follows.

Key Recommendations:

1. Continue refining the school quality framework through focus groups with stakeholders
2. Continue refining and developing measures aligned with the framework
3. Continue conducting computer-based surveys and piloting cellphone-based surveys
4. Build out full website, make it public, and connect it to district and city webpages
5. Conduct programming for teachers, parents, and the public around the new data available
6. Use new data, within the district central office and the school committee, to set a broader range of measurable aims for the schools

The Somerville Framework

Prior to our work in Somerville, our research team constructed a draft school quality framework rooted in national polling data, small-scale local surveys, and existing research. To adapt the framework for the context of Somerville, we conducted a series of focus groups with stakeholders—parents, educators, administrators, and community members.

¹ Our participants matched the racial/ethnic demography of the city and mirrored its economic diversity. The group also included nearly-equal numbers of men and women, as well as nearly-equal numbers of public school parents and non-parents. Our participants had a broad range of educational backgrounds—from high school completion to terminal graduate degrees. And we included several participants for whom English is not a first language.

Measuring School Quality Beyond Test Scores

Final Report

We met with members of the Somerville Family Learning Collaborative (SFLC) and the SFLC community liaisons for each school, the principals of all K-8 schools and the high school, all district administrators, and approximately 50 teachers.

Ultimately, we distilled the framework into five main categories:

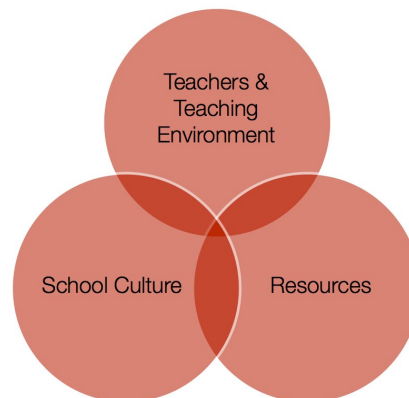
1. Teachers and the Teaching Environment
2. School Culture
3. Resources
4. Indicators of Academic Learning
5. Character and Wellbeing Outcomes

These categories, we believe, represent a coherent model of a good school, balancing inputs and outcomes.

The first three categories, broadly construed, are inputs. A school with a strong teaching environment, powerful school culture, and sufficient resources, is quite likely to produce positive outcomes. And a school lacking those elements will likely struggle.

Of course, these categories are not entirely independent of each other. Insofar as schools are ecosystems, each component effects the whole. Yet they are also independent enough from each other to merit separate categories.

A great teacher may choose to work in a school, for instance, even if its culture is problematic and its resources inadequate. Or a school may have a strong school culture, even if some of its teachers are weak and its resources limited. Still, the overlap among categories is worth noting because a good school must pay attention to all three. In other words, it is impossible to achieve ambitious goals in one category without targeting deficiencies in the others.

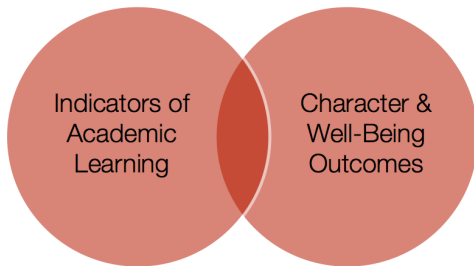


The other two categories in our framework—“Indicators of Academic “Learning” and “Character and Wellbeing Outcomes”—represent outcomes. As mentioned above, a school with strong inputs is quite likely to produce positive outcomes. Yet it is important to actually measure those outcomes. First, because it is conceivable, if improbable, that a school might earn high ratings for inputs but not outcomes. But second, and more likely, because a school may be succeeding more in one direction than another with regard to outcomes. For instance, a school may have great teachers, a positive culture, and adequate resources, and it may be channeling those strengths into academic learning alone, neglecting the character and wellbeing outcomes valued by parents, educators, and the public.

Measuring School Quality Beyond Test Scores

Final Report

As with our essential inputs, our key outcomes are not independent of each other. It is hard to imagine a school where students are developing high levels of civic engagement, strong work ethic, artistic and creative traits, and various health outcomes without actually learning anything.



This becomes particularly hard to imagine if the school possesses the essential inputs in our framework—inputs that include qualified and effective teachers, effective school leadership, and a strong academic orientation. Yet insofar as it is possible to succeed more in one domain than in the other, it is important to track both.

The categories in our framework were also balanced with regard to the exogenous (external to the school) and endogenous (internal to the school) nature of their influence. Good schools benefit from much that is outside of their control—financial support from the state, for instance, or high levels of parental engagement. And that should be recognized in any representation of school quality. Yet too much reliance on exogenous factors can send misleading messages about the work being done inside a school. Consequently, some balance is necessary to ensure that, in discussions about school improvement, stakeholders articulate well-rounded solutions that do not simply dump undue responsibility at the school’s doorstep.

Before presenting the framework in greater detail, one additional point is worth making: this framework should not be viewed as if it were carved in stone. Instead, it should be understood as a living document that can, and probably should, be updated and amended to match the values and concerns of stakeholders.

As can be seen in the framework (see next page), each of the five major categories is divided into several subcategories. Each subcategory, in turn, is composed of two or three separate measures. Consequently, each category is made up of six or eight separate measures—enough to ensure both comprehensiveness, as well as to ensure accuracy given the fact that multiple measures are less vulnerable to error than single measures. For much more detail, including the specific survey scales or data collected for each measure, see **Appendix A**.

Worth noting in the framework is the fact that four recommended measures—a portfolio measure of student achievement, a measure of problem solving, a measure of college performance, and a measure of creativity—were not used (they are denoted with double asterisks). These measures would not be impossible to produce, but they will require substantial effort. Challenging though they may be, we do feel that the district should pursue these in future work.

The Somerville Framework

ESSENTIAL INPUTS

1. Teachers And The Teaching Environment

1A. Knowledge and Skills of Teachers

- Measure 1A-i: Professional qualifications
- Measure 1A-ii: Effective practices
- Measure 1A-ii: Teacher temperament

1B. Teaching Environment

- Measure 1B-i: Teacher turnover
- Measure 1B-ii: Support for teaching dev. + growth
- Measure 1B-iii: Effective leadership

2. School Culture

2A. Safety

- Measure 2A-i: Student physical safety
- Measure 2A-ii: Bullying/Trust

2B. Relationships

- Measure 2B-i: Sense of belonging
- Measure 2B-ii: Student/teacher relationships

2C. Academic Orientation

- Measure 2C-i: Attendance and graduation
- Measure 2C-ii: Academic press

3. Resources

3A. Facilities and Personnel

- Measure 3A-i: Physical spaces and materials
- Measure 3A-ii: Content specialists + support staff

3B. Curricular Resources

- Measure 3B-i: Curricular strength and variety
- Measure 3B-ii: Class size

3C. Community Support

- Measure 3C-i: Family/school relationships
- Measure 3C-ii: Community involvement +
External partnerships

KEY OUTCOMES

4. Indicators Of Academic Learning

4A. Performance

- Measure 4A-i: Test score growth
- Measure 4A-ii: Portfolio/Alternative assessments

4B. Student Commitment to Learning

- Measure 4B-i: Engagement in school
- Measure 4B-ii: Valuing of learning

4C. Critical Thinking

- Measure 4C-i: Problem solving emphasis
- Measure 4C-ii: Problem solving skills

4D. College and Career Readiness

- Measure 4D-i: College-going
- Measure 4D-ii: College performance

5. Character And Wellbeing Outcomes

5A. Civic Engagement

- Measure 5A-i: Understanding others
- Measure 5A-ii: Appreciation for diversity

5B. Work Ethic

- Measure 5B-i: Perseverance and determination
- Measure 5B-ii: Growth mindset

5C. Artistic and Creative Traits

- Measure 5C-i: Participation in arts and literature
- Measure 5C-ii: Creativity

5D. Health

- Measure 5D-i: Social and emotional health
- Measure 5D-ii: Physical health

Additionally, it is worth mentioning that many stakeholders articulated particular concerns about serving the Special Education population in Somerville. Responding to their feedback, we developed a separate survey for teachers of Special Education students. For results from this survey, see **Appendix E**. In the future, however, district leaders and other stakeholders may wish to adapt the framework for other audiences, Special Education students in particular, as well as younger students (PK/K).

Recommendations:

1. Begin using the framework to structure district and school committee goals
2. Begin formally tracking data aligned with the framework
3. Introduce the framework to key stakeholder groups and conduct information sessions
4. Refine existing measures through data analysis and focus groups
5. Develop new measures to fill-out the framework

Data Collection and Interpretation

Many of the data points collected during this project were put together by the district. With the cooperation of Assistant Superintendent Vince McKay and the technical assistance of Kenya Avant-Ransome, we were able to draw on district-level information about factors like teacher turnover, student suspensions, and spending on professional development.

The majority of the new data collected were the result of a student survey and a teacher survey, administered electronically via SurveyMonkey (a “gold” level account was used for this study and proved sufficient). All students in grades 4 and up at each of the elementary schools completed these surveys. Students at the high school were scheduled to complete the survey but inclement weather and problems rescheduling prevented that from happening. All teachers in the district completed our survey.

Results from the teacher and student surveys can best be viewed through the custom-designed web tool built for this project. Below, however, are several key takeaways:

Key Teacher Takeaways:

1. The vast majority of teachers feel well-prepared and well supported
2. Teachers generally believe that professional development could be improved
3. Teachers generally have a high regard for their principals as leaders
4. Teachers are generally satisfied with resources but see particular issues worth addressing (class size, particularly, but also instructional support staff and access to materials)
5. Teachers generally believe the curriculum could be more well-rounded, with less focus on test-prep and more curricular diversity

Key Student Takeaways:

1. The vast majority of K-8 students find their teachers effective, supportive, and respectful
2. Students generally feel safe at school and have a strong sense of belonging
3. Students generally have positive relationships with adults and other students
4. Students generally value learning and take school seriously
5. Students generally express only moderate levels of excitement and joy in school

Once these data were collected, they were compiled anonymously by school and organized according to the logic of the school quality framework. The anonymous nature of the surveys, preserved by aggregation at the school level rather than the grade level, is a key aspect of eliciting honest responses, particularly from teachers, and we strongly recommend that this practice continue.

As alluded to earlier, we also conducted a separate survey among teachers of Special Education students. In order to protect their anonymity, we aggregated these data at the district level, since many schools have only a handful of Special Education teachers. Results from the Special Education survey can be found in **Appendix E**.

We also experimented with a parent survey. Research literature indicates that parent perspectives should be included when parents have first-hand information—about the behaviors and experiences of their children while at home, for instance, or about their own adult experiences in communicating with the schools. But second-hand information—about teacher quality, for instance—should not be included in any measure of school quality.

Our trial run sought to determine the feasibility of conducting a parent survey and results were not encouraging. Despite sending a note home to all parents in multiple languages, and conducting an online recruitment campaign, we had a very low response rate for our parent survey (particularly for our Spanish-language survey, despite work with the SFLC, and also among parents of high school students).

Results from the parent survey can be found in **Appendix F** (English-language survey) and **Appendix G** (Spanish-language survey).

Briefly, however, other interesting results include:

1. The vast majority of parents meet with teachers only a few times per year
2. The vast majority of parents are not involved with parent groups and do not volunteer
3. A majority of parents are pleased with the schools
4. A majority of parents are less than fully pleased with the range of school activities
5. Spanish-speaking families felt that respect for different cultures could be stronger

We might conclude, then, that parents are an untapped resource. And, though they are satisfied overall, that they desire more curricular and extra-curricular diversity in the schools. This is true for English- and Spanish-speaking parents responding to our survey.

Finally, in terms of data collection, we conducted a pilot study in which we utilized cellphones as survey collection devices. Though the data generated through our cellphone-based surveys were not used in any of our school quality portraits, we do believe that the method showed great promise. 53 teachers participated in our smartphone study and 36 of them responded to 98% or more of questions. Five additional teachers answered more than 50% of questions; and on the whole, 75% of questions asked were answered. As an incentive, teachers were paid \$25 to put the app on their phones and were promised an additional \$25-50 if they continued texting responses for the duration of the study. A much smaller percentage of students—roughly 80 students total—participated in our text messaging survey program. Yet that project, too, showed great promise. Participating students answered roughly 60% of questions, despite being given modest incentives (entrance into a drawing for cash prizes).

We believe that this method of data collection—through cellphone-based surveying—should be pursued, as it does not interfere with instructional time, and as it offers the potential to capture “real time” data. The cost of this method of data collection will only decrease over time.

Data Interpretation

After compilation of the data, we sought to interpret outcomes for users. If a school earned an average score of 3.8 on a survey question about student/teacher relationships, for instance, was that a success? A failure? Ultimately, this is a core aspect of presenting data to most audiences, as any non-expert will struggle to define success.

In order to answer questions such as these, we turned again to the community. Through a series of focus groups and surveys with the SFLC community liaisons, district administrators, principals, and teachers, we asked stakeholders to help us establish a “zone of approval” for each of the five major categories in the framework. To do this, we presented participants with fictional scenarios and asked them to identify those they deemed acceptable. This range then became our “zone of approval.”

After establishing the “zone of approval,” we sought to identify other zones by leveraging the concept of growth. That is, rather than saying a school is “failing,” we sought to frame a school’s current work as being within a certain number of years of its targets—an approach rooted in the belief that all schools can improve.

In order to do this, we asked focus group participants and survey respondents to tell us how much a school can reasonably grow in two years—a time period identified as “not too long to wait” by parents, and “sufficient time to improve” by educators. Being located in

the “growth zone,” consequently, indicates that a school can (with focus and support) reach its target goals within two years. We also used this two-year growth figure to establish a “watch zone” below the growth zone and, below the watch zone, a “warning zone.” For more information about these zones, see **Appendix B**.

Many of the data from our study lent themselves naturally to this approach. All of our survey measures, for instance, produced 1-5 Likert scores, which can be easily converted into percentages.² Many other data, however, were less straightforward to incorporate into this method of interpretation. For figures like teacher turnover rate or the ratio of counselors-to-students, our approach was to generate z-scores (a statistical measure of a score’s relationship to the mean for that figure) through the use of state or national averages. These z-scores were then scaled to match our Likert scores, with each number 1-5 representing a standard deviation below or above the mean.

While we are confident in the overall approach to data interpretation, we do recommend that more work be done to further refine these zones. Currently, they are relatively crude estimates, and additional focus groups and surveys—conducted by using the web tool developed through this project—would do a great deal to sharpen the edges of these zones. We were particularly unhappy with the need to rely on z-scores, which are inherently relativist in nature (in such a scoring system, a school cannot be “good” unless it is better than other schools). Again, further focus groups and surveys should be conducted to resolve this.

Recommendations:

1. Continue to conduct district-wide student and teacher surveys via SurveyMonkey
2. Maintain survey anonymity by aggregating data at school level rather than grade level
3. Consider ways to include parent perspectives
4. Conduct surveys and focus groups to refine outcome “zones”
5. Conduct surveys and focus groups to reduce reliance on z-scores
6. Scale-up the teacher smartphone survey pilot
7. Scale-up the student text messaging survey pilot

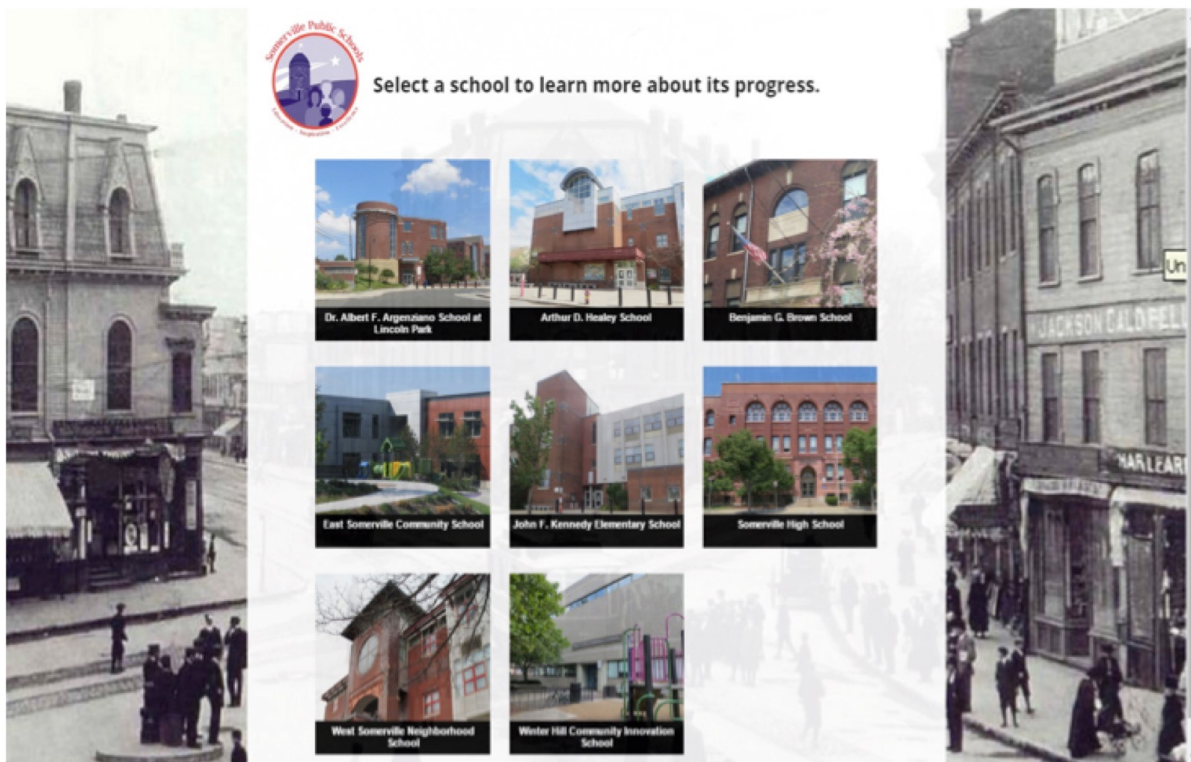
The Website

Working with two developers—Jared Cosulich and Alec Resnick—our team built an online data portal that visualizes the new data in a straightforward and user-friendly manner. This data can certainly be printed out for those lacking internet access. For an example of what can be done, see **Appendix F**. That said, the interactive nature of the tool is one of its major benefits.

² The following is an example of a Likert scale: *Strongly Disagree - Disagree - Undecided - Agree - Strongly Agree*

Measuring School Quality Beyond Test Scores Final Report

On the home page of the site (which is not yet online, and for which the district and the city will need to determine hosting, etc.), each school has its own icon:



Upon clicking that icon, users are taken to a school-specific page. We intentionally did this to prevent people from viewing schools as being in competition with one another. As we have maintained throughout this project, this tool is not a device for shopping or for pitting schools against each other. Insofar as it may be used by those outside the district office or the school committee, it was designed to empower people to make more informed decisions, to advocate for the schools, and to engage in a more informed manner with teachers, other parents, school administrators, and district leaders.

On the home page for each school, users see the five major categories of the Somerville framework, a description of each category, and a color-coded “data interpretation” for each category (green, light green, yellow, orange).

All of this is projected over an image of the school itself (See top of page 12 for graphic).

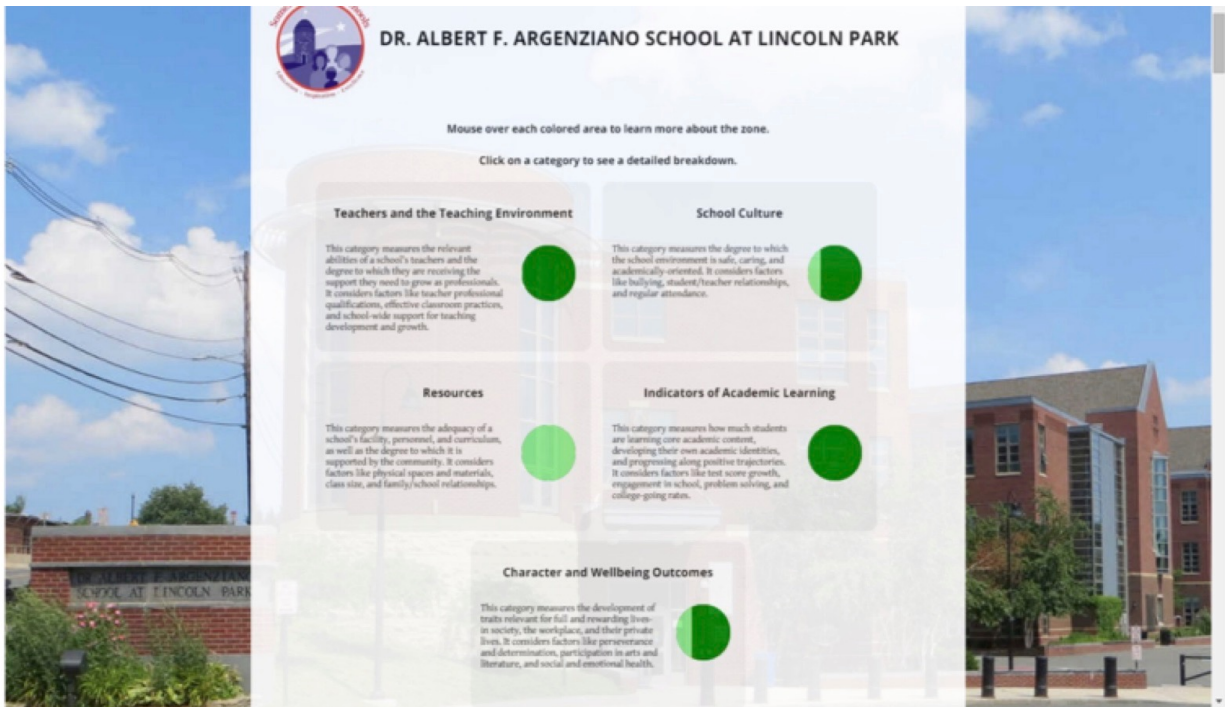
When users click on one of the categories, they are taken to a more detailed view—replete with subcategories, descriptions of subcategories, specific measures, and “visualized data” for each measure (See **Appendix H**).

Currently, users cannot click on measures to dive deeper into the data. Instead, they either see bar graphs representing the distribution of survey responses for a particular

Measuring School Quality Beyond Test Scores

Final Report

scale, or they see the school data alongside a benchmark figure (state average, national average, or expert recommendation).



Ideally, however, users would be able to click on each measure and see in even greater detail the results of data collection. For instance, users can currently see a bar graph aggregating all student responses to survey questions about student/teacher relationships. Yet students were asked a survey scale of 12 separate questions about their relationships with their teachers. Consequently, much of that data is being compressed in the current representation. We recommend additional web development so that users can see averages for each individual survey question. Responses to a question like "How much do you enjoy learning from this teacher?" may, in fact, be quite different from responses to "How friendly is this teacher toward you?" Allowing users to see individual questions, as well as to learn more about how data are collected, would be work worth investing in.

The web portal that we have constructed is an exemplar, representing a promising approach to visualizing data. Still, it is a prototype. This project was funded on a very modest budget and relied on hundreds of hours of donated time. Web development was no exception. If the school committee and the district are committed to better and more usable data, more investment in the website is necessary.

Yet while the website has greater potential, it also represents a major improvement over what is currently available. The results from our May 30 polling, sponsored by the SomerPromise office, indicate that residents of Somerville would be both more

Measuring School Quality Beyond Test Scores
Final Report

knowledgeable about and more confident in the public schools if they had access to our web tool.

During that poll, we created two groups—consisting of 22 and 23 participants, respectively—of demographically-representative participants. One group worked with existing data, available through the Department of Elementary and Secondary Education’s school profiles website. The other group worked with the web tool developed through our project—a tool that visualized the new data collected over the 2014-2015 school year. Participants were polled four times—upon arrival, after initial data viewing, after small group conversations, and after mixed-group conversations—about a range of questions. More data about this poll will be produced in the coming months.

While further data analysis will take time, three key findings did emerge from the poll.

First, the new data tool appears to have had *a positive influence on perception*, at least with regard to the schools users are *not familiar with*.

Participants gave similar ratings to the schools they expressed familiarity with, both across time—from first rating to third rating—and across the two different data sets. But when it came to the schools that people were not familiar with, the new data group gave *substantially higher ratings than the state data group did*. The vast majority of participants responded “I don’t know” when asked about the quality of their randomly-assigned schools. Consequently, it is impossible to gauge the difference between first rating and third rating. Nevertheless, the third ratings were markedly higher among users working with new data, as the chart below reveals.

Impact on Perception of School Quality

	Initial Rating		Third Rating	
	Familiar	Random	Familiar	Random
New Data Group	3.6	n/a*	3.4	3.5
State Data Group	3.8	n/a*	3.5	2.8

* No average score is available because the vast majority of users responded “I don’t know”

In short, it appears that the new data system gives users a perspective similar to that of those who have first-hand experience. The state data system, by contrast, appears to give a more negative impression of schools to those unfamiliar with them.

Second, it appears that the new data system may change the views of those who have never even used it.

Measuring School Quality Beyond Test Scores

Final Report

After viewing data on their own, participants were placed in mixed groups (new data viewers and state data viewers). The purpose of mixing groups was to see if engagement with data might affect *those who had not actually looked at it*. For instance, would talking with someone from the new data group change the opinions of a participant from the state data group, even if s/he had not interacted with the new data?

As the chart below indicates, these conversations did not substantially affect participants working with the new data. But notably, those working with state data had somewhat higher opinions of the schools *after talking with those who had viewed the new data*. This is surprising, because it reveals the possibility that the new data system might impact the views of people who never use it.

Impact on Perceptions of Those From Other Groups

	Third Rating		Final Rating	
	Familiar	Random	Familiar	Random
New Data Group	3.4	3.5	3.3 (-0.1)	3.4 (-0.1)
State Data Group	3.5	2.8	3.7 (-0.2)	3.1 (+0.3)

Word-of-mouth is historically one of the leading ways that parents and community members obtain information about school quality. Sometimes that word-of-mouth information is high quality—passed along by parents of current or former students, or by educators. But, as research has documented, such word-of-mouth is no less powerful even when it is drastically uninformed, as it can often be. In recent years, the use of standardized test scores and related accountability metrics to rate schools has helped structure these face-to-face conversations. Yet because test scores measure more about a school's demography than about its programming, they have done little to dispel overly-negative views of urban schools.

This experiment indicates that our new data system may have the power to counter misinformation. As can be seen above, users *working with state data had significantly higher perceptions of their randomly-assigned schools after talking with members of the other group*—those who had worked with the new data. This is a promising finding worth exploring further.

A third major finding worth discussing is the impact of the new data system on the degree to which stakeholders believe they can identify school strengths and weaknesses. As the chart below indicates, participants working with the new data had *substantially higher confidence in their knowledge about the schools*. This was particularly true with regard to the schools they were randomly assigned (rather than those with which they were familiar). Ratings among new data users *went up by 50%* between the initial rating and the final rating. And those final ratings were *nearly 50% greater than those of the state data group*.

Measuring School Quality Beyond Test Scores
Final Report

Ability to Identify School Strengths and Weaknesses

	Initial Rating		Final Rating	
	Familiar	Random	Familiar	Random
New Data Group	3.4	2.3**	3.9 (+0.5)	3.5 (+1.2)
State Data Group	3.4	1.8**	3.5 (+0.1)	2.4 (+0.6)

** 50% of respondents answered “I don’t know”

In sum, the new data system gives users a much more detailed view of the schools—to the extent that users can identify strengths and weaknesses at a randomly-assigned school almost as well as at a school with which they are familiar.

On the whole, we believe it is safe to say that use of the new data tool will not harm perceptions of school quality. In fact, they may improve the standing of the Somerville public schools. We also believe that use of the new data tool will make community members better informed about the schools, even if they are already familiar with them. And it may even improve the knowledge of users who never even access the new data.

While our poll dealt exclusively with parents and community members, it is also important to state that these groups are not the only potential users of this web tool. Teachers and principals might make very powerful and productive use of this tool—setting a broader range of goals, identifying more specific and measurable aims, communicating more effectively with the district about their needs and desires, and more transparently communicating with parents. We believe that professional development conducted around this web tool can empower educators, help them make context-specific decisions about their schools, and give them recognition for the full range of things that they are already doing. Similarly, district leaders and the Somerville school committee might use this tool to set and track goals across the city’s schools.

As a final note about the website, it is worth mentioning that very little work of this nature has been done anywhere in the nation. Consequently, we recommend that the district and the school committee keep track of what is working and what isn’t, so that the tool can continue to grow and evolve.

Recommendations:

1. Build-out website to full potential
2. Make the website public and promote it
3. Introduce website to teachers and principals
4. Keep track of what is working and what isn’t, for future refinement

The Schools in Somerville

Overall the new data produced by this project indicate that the public schools in Somerville are doing well. This should come as no surprise to anyone with experience in the schools. It is, however, worth noting that it may come as some surprise to those who rely on cruder measures for information about a school's quality—audiences like prospective parents or the state Department of Elementary and Secondary Education.

Still, determining school quality—something best done through direct interaction with a school—was not the intention of this project. Instead, the project sought to make conversations about school quality more consistent and more accurate, to enable a broader and more specific set of goals, to promote engagement among families and community members, to facilitate communication between various stakeholder groups, and to credit schools for the full range of things they do.

Nevertheless, a few observations about the schools in general merit mention here.

The first observation is that, despite increases in funding for public education in Somerville over the past several years, “Resources” is the lowest-rated category for almost all schools. This does not necessarily mean that more spending is necessary, as the rating was not tied to a dollar figure. But it does indicate that attention should be paid to the adequacy of facilities, personnel, curricular resources, and mechanisms for generating community support.

The second observation is that the schools appear to be meeting “Indicators of Academic Learning” goals more effectively than goals related “Character and Wellbeing Outcomes.” This makes sense. MCAS scores, whether through raw scores or Student Growth Percentile, provide clear and measurable benchmarks that the schools have been under intense state-driven pressure to meet. Nothing comparable has existed for character and wellbeing outcomes. With time and new tools, we imagine this will change.

Third, “School Culture” at all schools is positive but can improve. With a few (very positive) exceptions, the same is true of “Teachers and the Teaching Environment.” For each school, it is worth digging into these categories to see where particular strengths and weaknesses lie. And though it may be that district-wide policy or programming might be necessary to address shortcomings consistent across all schools, it is more likely that each school will need to develop its own set of goals related to its specific strengths and weaknesses.

Fourth, of the schools included in this study, the high school appears to require the most support. This does not mean that the high school is struggling. In fact, the high school has a number of key strengths upon which to build. The data from this study indicate a solid corps of teachers, effective leadership, a safe environment, and adequate facilities. But the data also indicate a need for more resources, particularly to alleviate undue burdens on the school related to low levels of home and community support. Additionally,

Measuring School Quality Beyond Test Scores

Final Report

it should be noted that the high school is doing exemplary work in engaging students, teaching them to value learning, and promoting test score growth. Again, however, the tasks set out for the high school are quite challenging, and it is clear that the school is currently succeeding in the areas where leaders are concentrating their attention. This should not come as a surprise. But the district should also work to make it possible for educators and administrators at the high school to expand the scope of their focus—targeting an even fuller range of outcomes.

For a more detailed breakdown of the data, we recommend using the web portal, which we are turning over electronically to the district and the school committee along with this report. Additional results—from our student and teacher surveys—can be found in **Appendix C** (student survey) and **Appendix D** (teacher survey).

Finally, three schools were not included in this study: Next Wave, Full Circle, and the Michael Capuano Early Childhood Center. In the future, district leaders may wish to study the degree to which the school quality framework and associated measures are appropriate for those schools and their stakeholders. In the case of the Capuano, the challenge will be greater because of the challenge of collecting student surveys. If district leaders wish to include the Capuano in future efforts, it will take some creative thinking.

Recommendations:

1. Focus on one or two of the lowest scoring categories or sub-categories for the district
2. Focus on the lowest scoring category or sub-category for each school
3. Think about ways to include Next Wave, Full Circle, and Capuano
4. Think about ways to include the perspectives of young students (below grade 4)

Moving Forward

The Somerville public schools are doing well and there is much to feel good about. The work of this project confirms that, and the products of this research can be used to empower all stakeholders in the community.

Work should not end here, however. Over the next five years, the school committee and district leaders should work to integrate this new data system into all aspects of public education in Somerville—inside the schools, in the community, within the district, and among policymakers.

That work will take time. And it will also take resources. But we believe that this project has an incredible upside for a relatively small commitment of time, energy, and funds.

Our team is willing to support the Somerville Public Schools in such work. Still, we are committed to phasing ourselves out. This is not only so that our team members can move on to other projects, but also because the final measure of success for this project will be a

Measuring School Quality Beyond Test Scores

Final Report

seamless integration into the district requiring no additional outside support. In line with that, we are planning a stepped-down level of involvement for the 2015-2016 school year, and will be working to build capacity with the district, the school committee, educators, and community members so that we can begin moving toward zero involvement within the next several years.

Acknowledgements

This project would not have been possible without the creative leadership of Superintendent Tony Pierantozzi and Mayor Joe Curtatone. In the district office, Vince McKay, Kenya Avant-Ransome, and other staff were incredibly supportive. Anna Fox Doherty and the SomerPromise office supported the evaluation component of this project. Sarah Davila, Meghan Bouchard, and the SFLC community liaisons were an invaluable resource. The principals of the schools and scores of educators gave their time generously and were always willing to share their insights. Alec Resnick and Jared Cosulich dedicated countless hours to technological support for this project, including the development of our text messaging program and the development of our web portal. Ben Keeler and the staff at EduMetrics developed our teacher smartphone app. And research assistants at the College of the Holy Cross, Harvard University, and Michigan State University played a key role in supporting the project.

List of Appendices

- Appendix A: Somerville School Quality Framework
- Appendix B: Outcome “Zones”
- Appendix C: Student Survey Results
- Appendix D: Teacher Survey Results
- Appendix E: Special Education Survey Results
- Appendix F: Parent Survey Results
- Appendix G: Parent Survey Results (Spanish)
- Appendix H: Website Examples

Measuring School Quality Beyond Test Scores

Year 2 Final Report

May 15th, 2016

Project Director:

Douglas Gagnon

University of New Hampshire

douglas.gagnon@unh.edu

Principal Investigator:

Jack Schneider

College of the Holy Cross

jschneid@holycross.edu

Executive Summary

In 2014, our research team contracted with the Somerville Public School to build a new, more complete measure of school quality. This project aims to better capture the work that schools do, and to present data in a clear manner to broad audience. A more holistic school quality framework was drafted using national polling data, a broad research base, and refined through gathering feedback from stakeholders, including teachers, parents, administrators, and community members. Once a framework was settled, the team set to determining how its constructs could be measured. Teacher and student surveys were drafted and issued to Somerville schools. In addition to these perceptual data, administrative data was also collected and incorporated into the framework. In order to communicate these assembled data, a user-friendly webtool was created. This tool includes a page for each school, representing data using simple to understand color-coded zones and textual descriptions. The Year 1 Report describes this work in detail, closing with a number of recommendations for future work; this Year 2 Report highlights the efforts that have been made to address these recommendations.

First, a comprehensive set of analyses of year 1 data was completed. We find a remarkably rates of survey completion, thus giving us confidence that their measures represent students and teachers in Somerville schools. A thorough examination of survey scales was conducted, including factor loading and reliability testing. This analysis identified numerous survey questions that were not performing optimally, and survey scales were edited and trimmed down as a result. Next, a correlational analysis was conducted for all metrics used in the framework. Moderately strong associations were found among the major categories within the framework, providing some evidence in support of the validity of this framework as a broader measure of school quality. In an effort to refine the benchmarks of these data—essentially, determining what constitutes a “good” score for a given school—a survey was created that aims to capture notions of quality among a broad audience. This survey is tentatively set to be administered next month.

In order to further validate the Framework with a more diverse group of community members, two additional focus groups were conducted in March 2016. These meetings, which were organized and facilitated with the help of the Somerville Family Learning Collaborative and The Welcome Project, elicited the feedback of over 30 adults whose home language is not English. All members of these focus groups, whose native languages include Spanish, Portuguese, Creole, and Arabic, have children in Somerville Public Schools. Based on their feedback of the Somerville School Quality Framework, a new question was added to the survey.

Another stream of work revolved around refining and improving the webtool. First, our research team facilitated the feedback of all Somerville principals through the review of their school’s data on the webtool. Concerns voiced by the principals—in terms of clarity, accuracy, ease of use, etc.—were used to make refinements to the webtool. Next, a series of webtool user tests were conducted using the newly refined version of the webtool. Opinions from a wide group of stakeholders were facilitated during these user tests. In addition, these meetings with principals largely confirmed the accuracy of the data presented in the webtool.

In total, the work completed in the past year has improved the accuracy, scope, and community support for the Somerville School Quality Framework. Additionally, important user-driven changes are being made to how school quality data are visualized via the webtool. Overall, we believe that the Somerville School Quality Framework and its associated tools could soon be made public. Furthermore, the framework could serve as an alternative model for school quality accountability.

School Quality Framework

In the first year of this project a framework for school quality was developed. This was done in close collaboration with teachers, administrators, and members of the community. A major goal of developing this framework was to include a wide range of measures that fully reflect the diffuse purposes of schools. The framework was structured such that it includes inputs—things that schools do—as well as outputs—evidence of success in the form of student outcomes. Ultimately the framework was distilled into five categories:

1. Teachers and the Teaching Environment
2. School Culture
3. Resources
4. Indicators of Academic Learning
5. Character and Wellbeing Outcomes

Each of the five major categories is divided into several subcategories. Each subcategory, in turn, is composed of two or three separate metrics. Consequently, each category is made up of six or eight separate metrics in an effort to ensure both comprehensiveness and accuracy. The complete framework, along with descriptions of each measure, may be found in the appendix.

These five categories remained intact after the second year of the project. However, there were refinements made to the metrics which constitute these categories; these refinements are detailed in the subsequent section.

Data Analysis

The work in year 2 of this project began by analyzing the data collected in year 1. First we examined completion trends for the surveys. Students in grades 4 and above at each of the elementary/middle schools were issued surveys, with students in grades 3 and below being excluded due concerns of age appropriateness regarding the comprehension required to complete the survey. This resulted in a student sample of 1607 students, or roughly 98 percent of non-excluded population. We judge completion rates by examining the proportion of students who answer the last question. Overall, we find that students generally persist in taking the survey. For instance, 93.4 percent of fourth graders finished the survey, while 96.6 percent of 8th graders did so. Thus, we are confident that the student survey was age appropriate, and that its results represent the entire population of students in the district. The teacher sample was issued and completed by 324 teachers in the district, which we take to be the universe of teachers.

The first major portion of this analysis involved reliability testing and scale loading of the survey-based metrics. Through this work we can make adjustments to survey scales in future administrations, hopefully improving reliability while at the same time reducing the amount of time required to complete the surveys. The second major portion included a series of correlational analyses. Through such descriptive procedures, we can get a better understanding of the associations between the measures that constitute our school quality framework, which in turn can be used to bolster claims of its validity. We discuss each of these major analytical threads below.

Survey Analyses. When measuring a construct using surveys (e.g. interest in students, or community engagement), multiple questions are used in order to increase reliability—and ultimately our confidence that what is being measured is the construct of interest and not noise, or measuring error. Methodologically speaking, each survey item (i.e., individual question) within a construct should be related, or positively correlated, to each other. To examine the degree of this internal consistency, we calculate Cronbach's alpha (α) for each construct, or survey metric. A rough rule of thumb is that α should be at least 0.7 in order to be considered a reliable metric. We also examine the relationship between each individual survey item and its corresponding scale. In general, we want the inclusion of items to improve (increase) α , and items found to decrease α are worthy of attention. Items that deserve especially

close scrutiny are those that not only decrease scale reliability, but also exhibit a negative correlation to the other items in the scale. Such items have the inverse relationship to the construct of interest than was theoretically posited.

We find that all survey-based metrics exhibit a reliability of 0.6 or greater, and most meet the 0.7 threshold. These are encouraging results. For those metrics with lower reliability, we revisited problem items. A few items were edited, and over a dozen items were removed completely. We believe that these edits to survey-based metrics should improve their reliability in subsequent survey administrations. In addition, one new survey item was added to bolster the range of one particular metric.

Correlational Analyses. The second broad category of analyses presented here explore the associations between components of the school quality framework. These correlational analyses are at the school level, and are conducted for all levels of measures in this project: metrics (lowest level) sub-measures, measures, and main categories (highest level of aggregation). In general, measures that share a common, broader construct should exhibit positive correlation; such a relationship helps to establish convergent validity. However, this is not unequivocally true, and there may be entirely compelling reasons to form a broader construct from domains that are inversely related. For instance, one might expect that schools with smaller class sizes (3Biia) also exhibit better perceptual data around class size (3Biib). However, we find that these two metrics actually exhibit a weak-to-moderate *negative* correlation. In other words, teachers in schools with better class sizes actually exhibit more displeasure with class size than do teachers in schools with worse class sizes. One possible explanation for this finding is that schools with small class sizes may serve more challenging populations, and thus teachers in these schools perceive a need for smaller class sizes despite the actual trends. Both metrics are supplying useful information, and one could argue for the inclusion of both metrics in forming the Class Size sub-measure (3Bii) despite this observed relationship. Thus, we chose to include both class size metrics. However, we do recommend dropping three initial metrics due to poor fit and general inappropriateness: percent of licensed teachers, average professional development spending, and support staff-to-student ratio. See appendix I for school-level correlations for all remaining metrics.

When measures are aggregated to the main category level, we find that all correlations between main categories of school quality are positive (see table 1). Such a result is theoretically appealing, as it seems likely that different measures of school quality should reinforce and promote each other. If one believes, for instance, that a strong school culture likely improves the teaching environment, and vice versa, then the results presented here should be encouraging.

Table 1
Correlations of Main Categories of School Quality

	Teachers and the Teaching Environment (1)	School Culture (2)	Resources (3)	Academic Learning (4)	Character and Wellbeing (5)
Teachers and the Teaching Environment (1)	1				
School Culture (2)	0.64	1			
Resources (3)	0.44	0.25	1		
Academic Learning (4)	0.29	0.70	0.42	1	
Character and Wellbeing (5)	0.66	0.49	0.25	0.29	1

Validating the Framework

The School Quality Framework was reviewed by a number of focus groups in 2014-2015. In order to validate the Framework with a more diverse group of community members, two additional focus groups were conducted in March 2016. These meetings, which were organized and facilitated with the help of the Somerville Family Learning Collaborative and The Welcome Project, elicited the feedback of over 30 adults whose home language is not English. All members of these focus groups, whose native languages include Spanish, Portuguese, Creole, and Arabic, have children in Somerville Public Schools.

These meetings largely reaffirmed the appropriateness of the school quality framework. Meeting participants emphasized the importance of factors such as community integration and school safety in their conception of school quality—factors which the current framework reflects. However, the groups also identified an aspect of community support which the framework was lacking, and for this reason a survey question was added to the appropriate scale.

Data review sessions were held using the webtool with each of the eight principals of traditional schools in the district. These sessions, which lasted roughly 45 minutes, involved the review of the student survey, teacher survey, and district administrative data which together comprised the metrics in the Somerville School Quality Framework. The primary purpose of these meetings was to verify the accuracy the data. In addition, principals were urged to comment on the validity of the Framework: Are measures properly benchmarked? Are measures complete? Does the Framework include measures that should not be included? Finally, principals commented on the usability and clarity of the data webtool (webtool refinement is discussed in the next section).

Principals in large part validated the accuracy of the school quality data. A few principals expressed some concerns that additional metrics should be added, or that teachers and students perceptions might be incomplete, but there was there was a very strong consensus that the perceptions presented were accurate. Principals also reported that, in general, the measures already included were appropriate. Very little feedback was offered as to the appropriateness of data benchmarking. In general, feedback on the Framework and webtool was quite positive.

The next important step in validating and improving the utility of the school quality framework was to create a benchmarking survey. The purpose of the survey is to gather stakeholder feedback on the appropriateness of selected school quality benchmarks for all metrics in the framework. It does so by presenting survey takers with a description of each metric, a description of how the proposed benchmark was developed (if applicable), and a range of possible metric values, and finally by facilitating perceptions as to the appropriateness of various possible metric values. The survey has been constructed and entered into Survey Monkey, and it is tentatively set to be issued in May or June. The Year 1 Report recommended conducting surveys to reduce the reliance on z-scores, and the administration of this survey is a step in this direction. By establishing criterion benchmarks for the metrics included in our school quality framework, we are not limited to making normative comparisons (or comparisons to the mean) for such metrics.

Webtool Refinement

In addition to gaining valuable insight into the accuracy and validity of data, the principal data review sessions also served as a means to assess the usability of the webtool. Feedback in this regard was generally quite positive, although a number of consistent suggestions were made by principals regarding how to improve the presentation of data in the webtool. For instance, principals voiced the need to include descriptions of each category, measure, sub-measure, and metric, perhaps through the use of a roll-over feature. Principals also felt that the source of each metric (student survey, teacher survey, or district records) should be clearly labeled. Whenever possible, consistent suggestions were used to refine the webtool.

In order to gain a wider perspective, user tests of the modified webtool were conducted on April 13th with various members of the Somerville community. The perceptions of users were captured and will be used to further refine and improve the webtool. This work identified bugs which need to be fixed, as well as design issues which require improvement. The design issues can be broken into two categories: things that are currently confusing to users, and things that would make the user experience easier. We are addressing the former before we address the latter. This work is being funded by a grant from the Spencer Foundation to Jack Schneider and the College of the Holy Cross.

Recommendations

The past year has seen important work in validating and improving a more holistic system by which school quality data may be collected and reported on in Somerville Public Schools. Moving forward, it is crucial that this momentum is maintained and that the continual adjustments are made. Along these lines we have a number of recommendations:

1. Complete a full set of diagnostic analyses on data collected in year 2.

The analyses presented in this Year 2 Report are an important step in validating the measures of school quality included in the School Quality Framework, but additional work should be conducted. Survey scales have been refined based on year 1 results, and therefore a fresh set of survey analyses are required with year 2 data. Additionally, a second year of data allow for numerous analytical threads which could further support the validity of the school quality framework.

2. Continue benchmarking efforts.

The tentatively scheduled benchmarking survey should be administered and fully analyzed. The results of this work will help support benchmarking decisions in an effort to create a criterion referenced framework of school quality measures.

3. Conduct an additional round of principal user tests using year 2 data.

Given the changes that resulted from this work since principals last viewed school data using the webtool, a second round of principal user tests should be scheduled using year 2 data.

4. Make the webtool live.

An investigation last year suggested that stakeholder knowledge and perceptions of school quality were increased after using the webtool. Given the tremendous progress has been made on the webtool since then, we believe that it will be even more effective in this manner.

5. Continue to work with under-represented groups.

In order for all community members to benefit from this work, it is important that continued efforts are made include under-represented groups. For instance, the survey and the webtool could be translated into other languages, and results from the special education teacher survey could be built into the webtool.

Acknowledgements

This work was done with the support of a number of district administrators, in particular Vince McKay and Kenya Avant-Ransome. Principals of Somerville Public Schools graciously offered their time and opinions in helping us refine what data should be collected and how it should be visualized. Nomi Davidson, Gillian Burleson, Tim Groves, and Meghan Bouchard were instrumental in organizing successful focus groups. Our webtool was improved markedly through the efforts of developer, Jared Cosulich. Finally, research assistants Victoria Gibson and Nick Parsons supported numerous facets of this work.

Appendices

Appendix I Somerville Framework for School Quality

ESSENTIAL INPUTS

1. Teachers and the Teaching Environment

This category measures the relevant abilities of a school's teachers and the degree to which they are receiving the support they need to grow as professionals. It considers factors like teacher professional qualifications, effective classroom practices, and school-wide support for teaching development and growth.

1A. Knowledge and Skills of Teachers

This subcategory seeks to determine the degree to which a school's teachers are prepared for their classroom assignments. It includes measures of teacher qualifications, effective classroom practices, and professional temperament.

Measure 1A-i: Professional Qualifications

This measure draws on anonymous teacher reports on their own comfort teaching grade-level, topics, and student body.

Measure 1A-ii: Effective Practices

This measure draws on anonymous student reports about factors like teacher clarity, support of students, and classroom management.

Measure 1A-iii: Teacher Temperament

This measure draws on anonymous student reports about the degree to which they perceive their teachers to be interested in and committed to them.

1B. Teaching Environment

This subcategory seeks to determine how well a school supports teachers and enables them to do their work. It includes measures of teacher satisfaction, effective leadership, and school-wide support for teacher development.

Measure 1B-i: Teacher Turnover

This measure draws on anonymous teacher reports on the percent of turnover in the teaching staff not due to retirement.

Measure 1B-ii: Support for Teaching Development and Growth

This measure draws on anonymous teacher reports on the quality of professional development.

Measure 1B-iii: Effective Leadership

This measure draws on anonymous teacher reports about the degree to which they trust their principals to make good school-wide decisions, as well as on the degree to which their principals are strong instructional leaders.

2. School Culture

This category measures the degree to which the school environment is safe, caring, and academically-oriented. It considers factors like bullying, student/teacher relationships, and regular attendance.

2A. Safety

This subcategory seeks to determine how safe the school environment is. It includes measures of physical safety, bullying, and trust.

Measure 2A-i: Student Physical Safety

This measure draws on data from the Youth Risk Behavior Survey on student safety, as well as on anonymous student reports about the degree to which they feel physically safe at school.

Measure 2A-ii: Bullying/Trust

This measure draws on anonymous student reports about the nature and frequency of school bullying, as well as on the degree to which students respect and get along with each other.

2B. Relationships

This subcategory seeks to determine how welcoming and caring the school environment is. It includes measures of student sense of belonging and of student/teacher relationships.

Measure 2B-i: Sense of Belonging

Measure 2B-ii: Student/Teacher Relationships

This measure draws on anonymous student reports about the degree to which they feel respected and cared for by their teachers.

2C. Academic Orientation

This subcategory seeks to determine the degree to which a school encourages students to focus on meeting academic challenges. It includes measures of student attendance and graduation, as well as of academic emphasis.

Measure 2C-i: Attendance and graduation

This measure includes the percentage of students chronically absent (more than 10% of days) from school and the percentage of students graduating on time.

Measure 2C-ii: Academic press

This measure draws on anonymous student reports about the degree to which teachers push them to do their best, work hard, and understand the material.

3. Resources

This category measures the adequacy of a school's facility, personnel, and curriculum, as well as the degree to which it is supported by the community. It considers factors like physical spaces and materials, class size, and family/school relationships.

3A. Facilities and Personnel

This subcategory seeks to determine the sufficiency of a school's staffing and facilities. It measures the quality of physical spaces and curricular materials, as well as the availability of content specialists and support personnel.

Measure 3A-i: Physical Spaces and Materials

This measure draws on anonymous teacher reports about their access to high-quality materials and facilities.

Measure 3A-ii: Content Specialists and Support Staff

This measure includes student-to-art-teacher and student-to-counselor ratios, and draws on anonymous teacher reports about the degree to which content specialists and support staff are available and effective.

3B. Curricular Resources

This subcategory seeks to determine the degree to which a school's classrooms include the essential resources teachers need. It includes measures of curriculum strength, curriculum variety, and class size.

Measure 3B-i: Curricular Strength and Variety

This measure includes the percentage of students completing the state core curriculum, the number of different classes offered per student, and the percentage of students participating in Advanced Placement courses in high school. It also draws on anonymous teacher reports on the strength and variety of the school curriculum.

Measure 3B-ii: Class Size

This measure includes the average class size at each school, and draws on anonymous teacher reports about the degree to which their classes are sufficiently small to support learning.

3C. Community Support

This subcategory seeks to determine the degree to which schools are supported by the surrounding community. It includes measures of family/school relationships, community involvement, and external partnerships.

Measure 3C-i: Family/School Relationships

This measure draws on anonymous teacher reports about parental engagement, as well as anonymous student reports about the degree to which their parents support them as learners.

Measure 3C-ii: Community Involvement + External Partnerships

This measure draws on anonymous teacher reports about the degree to which the school is an integrated part of the community.

KEY OUTCOMES

4. Indicators of Academic Learning

This category measures how much students are learning core academic content, developing their own academic identities, and progressing along positive trajectories. It considers factors like test score growth, engagement in school, problem solving, and college-going rates.

4A. Performance

This subcategory seeks to determine the degree to which students are learning core curricular content. It includes measures of growth on standardized tests and teacher perceptions of student academic growth; in the future, it will also include portfolio assessment scores.

Measure 4A-i: Test Score Growth

This measure includes school-wide scores for student growth on standardized tests, calculated by considering prior testing history and other factors.

Measure 4A-ii: Portfolio or Alternative Assessments

This measure draws on anonymous teacher reports about the efforts and abilities of their students. In the future this measure will also include expert evaluation of the work done by students in classrooms.

4B. Student Commitment to Learning

This subcategory seeks to determine the degree to which students are invested in the process of learning. It includes measures of student engagement and of how much students value learning.

Measure 4B-i: Engagement in School

This measure draws on anonymous student reports about their level of focus, participation, and interest in class.

Measure 4B-ii: Valuing of Learning

This measure draws on anonymous student reports about how important school is to them and how much they view themselves as learners.

4C. Critical Thinking

This subcategory seeks to determine whether students are learning to think critically about school subjects and the world around them. It includes measures of how much problem solving is emphasized in class and, in the future, will include a measure of student problem solving ability.

Measure 4C-i: Problem Solving Emphasis

This measure draws on anonymous teacher reports about how often their students have the opportunity to generative their own interpretations of material and apply knowledge to new situations.

Measure 4C-ii: Problem Solving Skills

In the future, this measure will include an assessment of student ability to address problems without obvious solutions.

4D. College and Career Readiness

This subcategory, applicable to high schools only, seeks to determine the degree to which students are prepared for college and beyond. It measures the percentage of students directly enrolling in two- or four-year colleges upon high school graduation and, in the future, will measure the college and career performance of high school graduates.

Measure 4D-i: College-Going

This measure includes the percentage of students enrolling in college immediately after high school graduation and, in the future, will include the college grades and employment status of graduates.

Measure 4D-ii: College Performance

In the future, this measure will include data on the percentage of students graduating from college in four or six years, as well as the percentage of students requiring college remediation.

5. Character and Wellbeing Outcomes

This category measures the development of traits relevant for full and rewarding lives—in society, the workplace, and their private lives. It considers factors like perseverance and determination, participation in arts and literature, and social and emotional health.

5A. Civic Engagement

This subcategory seeks to determine the degree to which students are prepared to thoughtfully meaningfully interact with others in a diverse society. It measures how well students understand the perspectives of others, as well as the degree to which they get along with those unlike themselves.

Measure 5A-i: Understanding Others

This measure draws on anonymous student reports about their ability to understand the views, emotions, and experiences of others.

Measure 5A-ii: Appreciation for Diversity

This measure draws on anonymous student reports about their level of comfort working with students from a wide variety of backgrounds.

5B. Work Ethic

This subcategory seeks to determine the degree to which students are willing to work hard on challenging tasks, even in light of setbacks. It includes measures of perseverance, determination, and orientation toward personal growth.

Measure 5B-i: Perseverance and Determination

This measure draws on anonymous student reports about their ability to pursue goals, work hard in spite of challenges, and finish what they start.

Measure 5B-ii: Growth Mindset

This measure draws on anonymous student reports about the degree to which they see themselves as capable of expanding their skills through hard work.

5C. Artistic and Creative Traits

This subcategory seeks to determine the degree to which students are being nurtured as artistic and creative people. It includes measures of student participation in arts and, in the future, will include a measure of student creativity.

Measure 5C-i: Participation in Arts and Literature

This measure draws on anonymous teacher reports about the frequency of student exposure to the arts.

Measure 5C-ii: Creativity

In the future, this measure will include an assessment of the ability of students to think outside-the-box when presented with different kinds of challenges.

5D. Health

This subcategory seeks to determine the health of students and the degree to which the school supports various health outcomes. It includes measures of student social, emotional, and physical health.

Measure 5D-i: Social and Emotional Health

This measure draws on anonymous student reports about how happy, calm, and focused they feel in school.

Measure 5D-ii: Physical Health

This measure draws on data from the Youth Risk Behavior survey on student physical health, as well as on anonymous teacher reports about student access to physical education and activity.

Appendix II
Correlations of School Quality Metrics (Within-Measure Correlations Highlighted)

	Professional Preparation Scale (1Aic)	Pedagogical Effectiveness Scale (1Aiiia)	Interest in Students Scale (1Aiiia)	Teacher Turnover (1Bia)	PD Scale (1Biib)	Teacher Principal Trust Scale (1Biiia)	Principal Instructional Leadership Scale (1Biib)	Student Safety Scale (2Aib)	Peer Victimization Scale (2Aiiib)	Peer Support Scale (2Aiiic)	Sense of Belonging Scale (2Bia)	Student Teacher Relationship Scale (2Biiia)	Chronic Absences (2Cia)	Academic Press Scale (2Ciiia)	Resources Scale (3Aia)	Art Classes per Student (3Aiiia)	Counselors per Students (3Aiiib)	Support Staff Scale (3Aiid)	Curricular Strength Scale (3Bif)	Class Size (3Biiia)	Class Size Scale (3Biib)	Parental Engagement Scale (3Cia)	Community Engagement Scale (3Ciiia)	State SGP Score (4Aia)	Student Achievement Scale (4Aiiia)	Student Engagement Scale (4Bia)	Valuing Learning Scale (4Biiia)	Problem Solving Scale (4Cia)	Appreciation for Diversity Scale (5Aiiia)	Grit Scale (5Bia)	Arts Exposure (5Cia)	Postive Affect Scale (5Dia)	Physical Activity (5Diiia)		
Professional Preparation Scale (1Aic)	1																																		
Pedagogical Effectiveness Scale (1Aiiia)	0.7	1																																	
Interest in Students Scale (1Aiiia)	-0.1	0.6	1																																
Teacher Turnover (1Bia)	0.4	-0	-0.3	1																															
PD Scale (1Biib)	-0.3	0.4	0.9	-0.4	1																														
Teacher Principal Trust Scale (1Biiia)	0.2	0.6	0.5	-0	0.7	1																													
Principal Instructional Leadership Scale (1Biib)	0.1	0.6	0.8	0.1	0.8	0.8	1																												
Student Safety Scale (2Aib)	0.8	0.3	-0.6	0.4	-0.7	-0.1	-0.4	1																											
Peer Victimization Scale (2Aiiib)	0.8	0.4	-0.3	0.7	-0.3	0.3	0.2	0.8	1																										
Peer Support Scale (2Aiiic)	0.6	0.5	0.2	0.6	0.2	0.7	0.6	0.3	0.7	1																									
Sense of Belonging Scale (2Bia)	0.2	0.6	0.5	0.1	0.7	1	0.8	-0.1	0.3	0.6	1																								
Student Teacher Relationship Scale (2Biiia)	-0.9	0.9	0.3	0.2	0	0.4	0.3	0.6	0.6	0.6	0.3	1																							
Chronic Absences (2Cia)	-0.6	0.2	-0.2	0.7	-0.2	0.2	0.1	0.5	0.7	0.8	0.2	-0.6	1																						
Academic Press Scale (2Ciiia)	-0.6	0.9	0.5	-0.2	0.2	0.4	0.3	0.4	0.3	0.2	0.4	0.8	-0	1																					
Resources Scale (3Aia)	-0.3	-0.2	0.3	-0.1	0.1	-0.3	-0.3	-0.4	-0.6	-0.3	-0.2	-0	0	-0.1	1																				
Art Classes per Student (3Aiiia)	-0.2	0.1	0.2	-0.3	0.5	0.5	0.6	-0.3	0	0.1	0.5	-0.3	-0.5	0.1	-0.7	1																			
Counselors per Students (3Aiiib)	0.5	0.6	-0	-0.3	0.1	0.4	0.2	0.4	0.4	0.1	0.3	0.4	-0.2	0.6	-0.7	0.6	1																		
Support Staff Scale (3Aiid)	0.2	0.8	0.9	-0	0.7	0.6	0.8	-0.2	0	0.4	0.6	0.6	0	0.7	0.1	0.2	0.2	1																	
Curricular Strength Scale (3Bif)	0.6	0.5	0.2	0.7	0.2	0.7	0.6	0.3	0.8	1	0.7	0.6	0.8	0.2	-0.4	0.2	0.1	0.4	1																
Class Size (3Biiia)	-0.6	-0.1	0.4	-0.6	0.7	0.3	0.5	-0.7	-0.5	-0.2	0.2	-0.5	-0.7	-0.1	-0.3	0.8	0.3	0.2	-0.2	1															
Class Size Scale (3Biib)	0	0.2	0.1	-0.2	-0.2	-0.4	-0.1	0.1	-0.2	-0.5	-0.4	0	-0.6	0.4	-0.3	0.2	0.4	0.2	-0.4	0.3	1														
Parental Engagement Scale (3Cia)	0.8	0.4	-0.2	0.5	-0.5	-0.2	-0.1	0.8	0.7	0.4	-0.3	0.7	0.5	0.4	-0.3	-0.3	0.3	0.1	0.4	-0.6	0.3	1													
Community Engagement Scale (3Ciiia)	0.7	0.3	-0	0.5	-0.4	-0.3	-0.1	0.6	0.5	0.4	-0.4	0.6	0.6	0.2	-0	-0.5	-0	0.1	0.3	-0.6	0.2	0.9	1												
State SGP Score (4Aia)	0.4	0	-0.5	-0.2	-0.2	0.1	-0.2	0.4	0.4	0.2	0	0.1	0.1	-0	-0.7	0.3	0.7	-0.5	0.1	0.1	-0.1	0.2	0	1											
Student Achievement Scale (4Aiiia)	0.9	0.6	-0.1	0.6	-0.1	0.4	0.3	0.7	1	0.8	0.4	0.8	0.7	0.5	-0.6	0.1	0.5	0.3	0.8	-0.4	-0.1	0.7	0.6	0.4	1										
Student Engagement Scale (4Bia)	-0.1	0.7	0.9	-0.4	0.9	0.7	0.8	-0.5	-0.2	0.2	0.7	0.3	-0.2	0.6	0.1	0.5	0.2	0.9	0.2	0.5	0.1	-0.3	-0.3	-0.3	0	1									
Valuing Learning Scale (4Biiia)	-0.1	0.5	0.8	-0.3	0.9	0.8	0.7	-0.5	-0.2	0.3	0.9	0.2	-0.1	0.4	0.1	0.4	0.1	0.7	0.3	0.4	-0.3	-0.5	-0.5	-0.3	-0	0.9	1								
Problem Solving Scale (4Cia)	1	0.6	-0.1	0.6	-0.3	0.2	0.2	0.8	0.9	0.7	0.2	0.8	0.7	0.4	-0.4	-0.1	0.4	0.2	0.7	-0.5	0	0.9	0.8	0.3	1	-0.1	-0.2	1							
Appreciation for Diversity Scale (5Aiiia)	0.6	0.5	-0.2	-0.3	-0.2	0.2	-0.3	0.6	0.3	0	0.2	0.5	0.1	0.6	-0.2	-0.1	0.7	-0.1	0	-0.3	0	0.2	0	0.5	0.4	-0	0.1	0.3	1						
Grit Scale (5Bia)	-0.2	0.3	0.6	-0.5	0.8	0.7	0.4	-0.5	-0.3	0.1	0.7	0.1	-0.1	0.2	0.3	0.3	0.1	0.4	0.1	0.3	-0.5	-0.7	-0.6	-0.1	-0.2	0.7	0.9	-0.4	0.2	1					
Arts Exposure (5Cia)	-0.3	0.5	0.7	0.2	0.2	0	0.4	-0.1	0	0.4	-0	0.6	0.3	0.3	0.4	-0.4	-0.3	0.7	0.3	-0.3	0.1	0.5	0.7	-0.6	0.2	0.4	0.1	0.4	-0.3	-0.1	1				
Postive Affect Scale (5Dia)	0.1	0.6	0.6	-0	0.8	1	0.8	-0.2	0.2	0.5	1	0.4	0.1	0.5	-0.2	0.5	0.3	0.7	0.6	0.3	-0.2	-0.2	-0.4	-0.1	0.3	0.8	0.9	0.1	0.2	0.7	0.1	1			
Physical Activity (5Diiia)	0.3	0.4	0.4	0.7	0.3	0.5	0.7	0	0.5	0.9	0.6	0.5	0.7	0.1	-0	-0	-0.3	0.5	0.9	-0.3	-0.4	0.3	0.3	-0.3	0.6	0.3	0.3	0.5	-0.3	0.1	0.5	0.5	1		