AASA JOURNAL OF SCHOLARSHIP & PRACTICE

AASA
THE SCHOOL SUPERINTENDENTS ASSOCIATION
WWW.AASA.ORG

Research and Evidence-Based Practice Advancing the Profession of Education Administration

Spring Volume 17 No. 1

Table of Contents

Editorial Review Board

Sponsorship and Appreciation

The *AASA Journal of Scholarship and Practice* would like to thank AASA, The School Superintendents Association, and in particular AASA's Leadership Development, for its ongoing sponsorship of the *Journal*.

We also offer special thanks to Kenneth Mitchell, Manhattanville College, for his efforts in selecting the articles that comprise this professional education journal and lending sound editorial comments to each volume.

The unique relationship between research and practice is appreciated, recognizing the mutual benefit to those educators who conduct the research and seek out evidence-based practice and those educators whose responsibility it is to carry out the mission of school districts in the education of children.

Without the support of AASA and Kenneth Mitchell, the *AASA Journal of Scholarship and Practice* would not be possible.

Assessing Responsibly

Ken Mitchell, EdD
Editor
AASA Journal of Scholarship and Practice

Since the 1983 *A Nation at Risk* report and the *No Child Left Behind* federal legislation in 2002, schools have been awash in a flood of testing and accountability regimens for the intended purposes of improving both student and educator success. However, during this extensive period of assessment-driven reform, measures of student, teacher, and school performance, often designed for administrative efficiency, have been scrutinized by researchers for their accuracy, questioned by practitioners for their usefulness, and challenged by educational leaders with a broader pedagogical vision of assessment than those of policy makers, legislators, and entrepreneurs with their special agendas.

Some of the lessons learned—and frequently ignored, even by educators—have been that assessment narrowly designed for efficiency and accountability has led to a corruption of its essential purpose, which is to generate an understanding of the needs of the learner and their learning environments. Such a design has also led to unintended consequences: Zhao (2018) warns of the "side effects" linked to test-based accountability systems: curriculum narrowing, distortion of instruction, exclusion of learners with certain "profiles," cheating, and the demoralization of learners and their teachers.

Zhao calls for researchers and research publications to assess both the main and side effects of educational practices before they are put into use, including assessment. In recent years educators, along with increasingly informed parents, and supported by researchers, have raised questions about the costs and benefits of today's performance measures. There have been renewed efforts to ensure that student assessment is implemented with fairness and equity, balanced in scope, volume, and frequency, and designed for accuracy and developmental appropriateness, culminating in the responsible and cautious, and, when necessary, skeptical, use of any data.

In recent years the *AASA Journal of Scholarship & Practice* has published research and essays on various and sometimes controversial assessment topics: Teacher, principal, and superintendent evaluation; the use of data to improve systems and achieve equity for students; test-based versus observation-based evaluation; the "gaming" of evaluation; the relationship between Common Core standards and lack of higher-order assessment task; and the design of policy-driven accountability reforms. This focus continues in our Spring 2020 issue, in which we examine the topics of assessment literacy, data assessment for school-wide improvement, and the effectiveness of value-added measures (VAM), a tool used by many states and districts for evaluating teacher effectiveness according to student performance data.

Thomas Guskey, in **"The Dark Side of Assessment Literacy: Avoiding the Perils of Accountability,"** describes the potential gains from enhancing assessment literacy but with a caveat, "Ideally it will broaden teachers' understanding of how to construct authentic assessments that tap student's performance in real-world contexts. It will help teachers design assessments that yield reliable results and are well-aligned with high level, cognitively complex student learning goals. Teachers will also know better how to gain valuable evidence from demonstrations, performances, projects, exhibits, and digital portfolios that can be used to guide improvements in instruction and student learning."

Yet, he warns that " … in the context of high-stakes accountability, where assessment-based decisions have serious and sometimes irreversible impact on the lives of students and their teachers both during school and afterward, increased assessment literacy also may lead teachers on a very different path. It may help them target their instruction and classroom assessments even more specifically on test preparation tasks."

Authors Martin-Kniepp and Lane, in "**Using Parallel Surveys and Reflective Conversations to Tap Perspectives and Promote improvement,"** take readers through an examination of how leaders and teachers can use data to improve the learning through a complementary use of parallel surveys: "Given the overwhelming amount of data principals have access to, and the fact that data sources are seldom integrated into accessible reports, principals could benefit greatly from formal opportunities to explicitly assess the data they have; ideally facilitated by external providers."

In the final piece Beardsley (**"All Value-Added Models (VAMs) Are Wrong, but Sometimes They May Be Useful"),** using a study of elementary students in a large suburban districts warns, "Findings indicate that ratings significantly and substantively differed depending upon the methodological approach used. Findings, accordingly, bring into question the validity of the inferences based on such estimates, especially when high-stakes decisions are made about teachers as based on estimates measured via different, albeit popular methods across different school districts and states.

Acknowledging that VAM is being used throughout the nation's schools, Beardsley warns of its limits and the importance of educators' awareness of these: "While the data produced by VAMs might be statistically sophisticated, contextual factors will *always* affect how VAMs play out in practice; hence, school administrators and teachers should be armed with as much knowledge as possible about when, why, and how VAMs should be used."

How educational systems design, implement, and use assessment, then interpret and apply the results can be life altering for students, as well as those educating them. While the issue's assessment perspectives may seem disparate, the authors' sentiments converge in a common quest for accuracy that is better achieved through sensitivity to context and an awareness of the potential consequences.

References

Zhao, Y. (2018). *What works may hurt: Side effects in education*. New York, N.Y.: Teachers College Press.

The Dark Side of Assessment Literacy: Avoiding the Perils of Accountability

Thomas R. Guskey, PhD
Senior Research Scholar
College of Education and Human Development
University of Louisville
Lexington, KY

Abstract

Educational measurement and evaluation experts generally agree that increasing stakeholders' assessment literacy will yield a variety of positive benefits, especially broadening the range of assessment formats teachers use to measure students' mastery of high level, more cognitively complex learning outcomes. But in the context of education accountability as currently structured in American schools, such efforts also may lead teachers to become more sophisticated in test preparation activities and to narrow both their instruction and classroom assessment practices specifically to enhance students' performance on prescribed, annual high-stakes accountability assessments. This article explains why that is so, describes the process by which it occurred in one state, and offers specific suggestions as to how it might be avoided.

Key Words

assessment, assessment literacy, accountability, classroom assessment, high-stakes assessment, matrix sampling, teacher attitudes, teacher commitment.

For nearly three decades, prominent experts in educational measurement have stressed the importance of assessment literacy (Popham, 2006, 2009, 2011; Stiggins, 1991, 1995; Xu & Brown, 2016). Some argue it may be the single most cost-effective way to improve our schools (Popham, 2018a). Assessment literacy is generally thought of as "the knowledge about how to assess what students know and can do, interpret the results of these assessments, and apply these results to improve student learning and program effectiveness" (Webb, 2002, p. 1). More recently Popham (2018b) described it as simply "an individual's understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions." (p. 2).

Improving assessment literacy could yield numerous positive benefits. It could broaden the ways teachers gather information on student learning and use that information to improve instruction. It could enhance students' use of assessments so they become more effective learners. It might even expand parents', families', and community members' interpretations of assessment results and encourage greater involvement in education endeavors.

Clearly the more stakeholders know about assessment techniques, interpretation, and use in decision-making, the better will be the educational decisions they make based on assessment results.

Education accountability systems as they are currently structured in the U.S., however, cast assessment literacy in an entirely different light. In the context of high-stakes accountability, increasing educators' assessment literacy could serve an unintended and far a more disconcerting purpose. This article explains that troubling purpose, why it is

likely, and what education leaders must do to avoid it.

Structure of Accountability Systems

Accountability systems in the U.S. emerged from increasing political involvement in education. They began with the *No Child Left Behind Act* (U.S. Congress, 2001) that made educators accountable to the general public for specific student achievement outcomes (Anderson, 2005).

Early accountability systems focused primarily on annual measures of student achievement in language arts and mathematic gathered in grades 3 through 8 and one year beyond. As these systems evolved, they expanded to include achievement in science and social studies, and took into account other measures such as attendance, promotion/retention rates, and graduation/dropout rates.

They further required that results be disaggregated to show progress among different subgroups of students (i.e., economically disadvantaged, English learners, ethnic or racial minorities, and students with disabilities) and to confirm reductions in achievement gaps. The *Every Student Succeeds Act* (U.S. Congress, 2015) has preserved annual grade-level testing but is less prescriptive about how the results are used in accountability systems.

The main challenge in modern accountability systems, of course, is how to accurately and reliably measure these student learning outcomes. Policy-makers and legislators typically pose the additional requirements on accountability systems that assessments of student learning not be too costly and be administered and scored efficiently so they do not require inordinate amounts of students' time.

Development of Accountability Measures

States varied in their approach to measuring these student learning outcomes. Most relied on external vendors to develop their assessments, trusting these vendors to ensure the assessments were aligned with the state's standards for student learning (Polikoff, Porter, & Smithson, 2011). Kentucky led the way in these efforts, establishing a statewide assessment and accountability system designed by experienced practitioners and several top experts in educational assessment (see Guskey, 1994).

A central feature of the Kentucky assessment program, known as the Kentucky Instructional Results Information System (KIRIS), was "on demand" performance events designed to assess students' higher level cognitive skills in several subject areas. These performance events required students to work together in teams to explain phenomenon or to find solutions to complex problems.

For each performance event, a small group of three or four students from a class or grade level was selected to engage in the event. Students worked on the tasks as a group but then prepared individual, written responses to specific questions or prompts regarding the event. Each student completed four events in the areas of math, science, and social studies. Some events were made interdisciplinary, however, combining science and math or math and social studies.

For example, a group of four students might be asked to observe and record data measuring the distance balls made of different materials bounce when dropped from a specific height. Based on their observations, the group would produce specified data tables or other products. From this information, each student was then asked to answer questions

individually that would depend on how well the group worked together to make the observations and record the data (Trimble, 1994).

Matrix Sampling

Research at that time showed that to get an accurate depiction of students' achievement of higher level cognitive skills in science or other subjects requires completion of 10 to 12 well-constructed performance tasks (Shavelson, Baxter, & Pine, 1991, 1992; Dunbar, Koretz, & Hoover, 1991; Messick, 1992). If each task in science took just ten minutes for students to complete, that would require two hours of testing time in science alone. Therefore, to economize the assessment process, the decision was made to use a strategy of "matrix sampling" for the performance events.

In matrix sampling, a substantial number of exemplary performance events, typically 12 or more, are designed for each grade level. Groups of three or four students randomly selected from each class or grade level complete four of the events, with each group completing different events Although no student completed every event, this allowed all events to be completed by some students at each grade level and all students to be involved in the assessment.

Results yielded fairly accurate and reliable estimates of students' achievement of higher level skills in science *at the school level*. If tasks and prompts from each event were well calibrated and reasonable numbers of students in various subgroups (i.e., ten or more) at each level completed events, it also permitted disaggregation of results for meaningful comparisons among student subgroups. Furthermore, because each student completed only four events, testing time in science was drastically reduced. But because each student completed only a limited number of events,

scores were not reliable at the individual student level; *only* at the school level. Since accountability focused on the school level, however, this issue was of little consequence.

Commitment of Teachers
Teachers want their students to succeed in school and to be confident in themselves as learners. They also want to feel they can influence students' learning and contribute to that success. These aspirations extend to students' performance on assessments that are part of accountability systems. Because of the important consequences attached to results from these assessments for students, for their families, for school leaders, and for the teachers themselves, students' performance on these assessments typically becomes a vital concern.

The Kentucky Instructional Results Information System (KIRIS) was clearly high-stakes for schools, school leaders, and teachers. It included financial rewards for schools that showed improved results and sanctions for schools that were not improving. State officials encouraged schools to provide teachers with the training necessary to prepare students for the new challenges of these performance-based assessments in science and other subjects.

Policy with Consequences Drives Practice
The effects on teachers' instructional activities of attaching high-stakes consequences to the results of performance assessments in science were profound. Not only did teachers begin to allocate more time to science lessons, they altered the way they taught science and the way they measured student learning on classroom assessments. Science lessons at all levels included more experiments and lab projects, and assessments involved data summary and interpretation, often integrating mathematics skills (Oldham, 1994).

The pressure for improvement in scores prompted many schools to devise professional development programs focused on the assessment formats and scoring procedures included in the accountability program (Cody & Guskey, 1997). A Rand investigation showed, for example, that all surveyed principals reported encouraging teachers to use materials specifically designed to guide students in inquiry-based events (Koretz, Barron, Mitchel, & Stecher, 1996). As a result, teachers included more performance tasks and authentic experiments as part of their instruction in science. They also taught students strategies for adapting their reporting based on specific scoring rubrics (Guskey & Oldham, 1996).

Funding Drives Policy
Unfortunately, these changes in teachers' instructional practices were short-lived. A newly elected group of state legislators who did not fully understand the matrix sampling procedures and were not particularly assessment literate raised concerns about assessment costs. Developing and piloting the performance events was costly. Scoring students' written responses to the science performance tasks was both time-consuming and expensive. In addition, although accountability remained focused at the school level, these legislators were concerned about the lack of reliability of scores at the individual student level.

Their response to these concerns was to impose drastic changes in the science assessments. Specifically, they wanted the assessments to require less time to administer and score in order to reduce the per-student costs. In addition, they wanted the assessment program to yield reliable data at the individual student level rather than just the school level.

Meeting these demands from legislators left the educational measurement experts who directed KIRIS with few options. The performance events were eliminated from the science assessments, as were the portfolios of student work that had been a foundational component of the language arts assessments. The statewide accountability assessments were returned to a more limited response format consisting of mostly multiple-choice items with a few extended-response items in each subject area.

The response of teachers to these changes in assessment format was predictable and immediate. Wanting to ensure their students did well on the new, restricted-response format science assessments, teachers revised their classroom assessments to more closely parallel the state assessments in science. Instructional strategies that resembled the performance events were abandoned in favor of activities and practices that prepared students for the more limited response format of multiple-choice items and brief, extended-response items.

As numerous studies have shown, teachers focus on the content tested and the way it is tested (Herman, 2004; Herman & Linn 2014). Arguments posed by state leaders in science education that students would do well on these restricted-response assessments when taught through a more inquiry-based approach to science fell on deaf ears. The teachers felt compelled to prepare their students for precisely what they would be asked to do on the new restricted-response, accountability assessments.

New Focus on Assessment Literacy

So what will result today from increasing stakeholders' assessment literacy? Ideally it will broaden teachers' understanding of how to construct authentic assessments that tap student's performance in real-world contexts. It will help teachers design assessments that yield reliable results and are well-aligned with high level, cognitively complex student learning goals. Teachers will also know better how to gain valuable evidence from demonstrations, performances, projects, exhibits, and digital portfolios that can be used to guide improvements in instruction and student learning.

Increasing students' assessment literacy will improve their use of assessment results to guide the correction of learning errors and help them become better managers and self-regulators of their own learning. Enhancing the assessment literacy of parents, families, and community members will inform their interpretations of assessment results. They will better understand what assessment results mean and the limitations of those results when drawing conclusions about the quality of instructional programs and schools.

But in the context of high-stakes accountability, where assessment-based decisions have serious and sometimes irreversible impact on the lives of students and their teachers both during school and afterward, increased assessment literacy also may lead teachers on a very different path. It may help them target their instruction and classroom assessments even more specifically on test preparation tasks.

Instead of broadening the array of assessment formats they employ, it actually may narrow what they teach, how they teach, and how they assess student learning to align more directly with the content and processes of those high-stakes assessments. It may make them even more highly skilled at focusing their instruction and classroom assessments on ways to improve students' performance on the limited but less expensive assessment formats

that provide the foundation for many of today's education accountability systems. And teachers will do this for noble reasons: because they care about the consequences attached to performance on those high-stakes assessments for their students, for them as teachers, and for their schools.

The Solution

This is not to suggest that efforts to improve the assessment literacy of all stakeholders should be abandoned. Teachers especially need help to broaden the ways they gather information on student learning and use that information to design effective instructional activities. They also need guidance in how to involve students in the assessment process so that students become insightful judges of their own performance and better self-regulators of their learning progress.

To avoid the unintended and potentially negative consequences that might accompany these efforts to improve assessment literacy, however, we must do two things. First, we must focus increased attention on perhaps the most influential but often most neglected group of stakeholders: policy-makers and legislators (see White, 2018). School leaders at all levels must make efforts to help these important decision-makers become more literate in every aspect of the assessment process.

In particular, policy makers and legislators need to understand that accountability assessments should model the types of assessment formats we hope teachers will use in their classrooms both to measure student achievement and to guide improvements in teaching and learning. In this way, teachers can teach to tests that are truly worth teaching to, and test preparation becomes a valuable instructional practice.

Credible high-stakes accountability assessments should focus on important 21st century learning goals, such as solving complex problems, reasoning and applying what is learned in new and different situations, communicating effectively, working collaboratively with classmates, and using higher cognitive processes. The best accountability assessments will also reflect authentic tasks and real-world contexts.

Assessments composed of multiple-choice and short, extended-response items certainly have their place and purpose. They offer an efficient and relatively inexpensive way to gather information about an important but fairly narrow range of student learning outcomes. Nevertheless, their limitations in measuring complex reasoning, communication, creativity, problem-solving, and other important learning goals must also be recognized.

Second, we must ensure the development of high-stakes accountability assessments is guided by valued learning goals rather than simply efficiency and cost. Cheap tests that don't measure the right things will not help us improve education. They are a waste of time and money, and a disservice both to educators and the students they teach. Increasing stakeholders' knowledge of the most valid means of capturing evidence on students' achievement of important 21st century learning goals will lead to more purposeful accountability assessments.

The Partnership for Assessment of Readiness for College and Careers (PARCC) assessments are a positive step in that direction. Although developing, administering and scoring these types of assessments will be somewhat more costly, the payoffs in terms of

students better prepared for success in school and beyond are vitally important.

With greater assessment literacy, policy-makers and legislators can demand better quality products from the vendors they hire to develop their state's accountability assessments.

They will understand the diverse assessment formats this requires, particularly performance events, projects, demonstrations, and portfolios of students' work. They also will understand the difference between reliability at the school level versus the individual student level, and know how school level reliability opens up a broader range of authentic assessment formats that can be employed with reasonable cost.

Increasing assessment literacy among stakeholders in the assessment process will help improve our schools, but only if efforts also target the policy-makers and legislators who make the important decisions about the format and structure of high-stakes accountability assessments.

Author Biography

Thomas Guskey is a senior research scholar in the University of Louisville's College of Education and Human Development. He is Professor Emeritus at the University of Kentucky. E-mail: guskey@uky.edu

References

Anderson, J. A. (2005). *Accountability in education*. Paris, France: International Academy of Education and International Institute for Educational Planning (United Nations Educational, Scientific and Cultural Organization).

Cody, C. B., & Guskey, T. R. (1997). Professional development. In J. C. Lindle, J. M. Petrosko, & R. S. Pankratz (Eds.), *1996 Review of Research on the Kentucky Education Reform Act* (pp. 191-209). Frankfort, KY: The Kentucky Institute for Education Research.

Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education*, *4*(4), 289-303.

Guskey, T. R. (Ed.) (1994). *High Stakes Performance Assessment: Perspectives on Kentucky's Educational Reform.* Thousand Oaks, CA: Corwin Press.

Guskey, T. R. (2008, March 16). Selecting school tests a matter of competence. *The Lexington Herald-Leader*, p. D-1 & 6. Available at: http://www.kentucky.com/589/story/348307.html

Guskey, T. R., & Oldham, B. R. (1996). *Despite the best intentions: Inconsistencies among components of Kentucky's systemic reform*. Paper presented at the annual meeting of the American Educational Research Association, New York City, NY.

Herman, J. (2004). The effects of testing on instruction. In S. H. Fuhrman & R. F. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 141-166). New York, NY: Teachers College Press.

Herman, J., & Linn, R. (2014). New assessments, new rigor. *Educational Leadership*, *71*(6), 34-37.

Koretz, D. M., Barron, S., Mitchell, K. J., & Stecher, B. M. (1996). *Perceived effects of the Kentucky Instructional Results Information System (KIRIS).* Santa Monica, CA: Rand Corporation.

Messick, S. (1992, June). The interplay of evidence and consequences in the validation of performance assessments. *ETS Research Report Series*, *1992*(1), i-42.

Oldham, B. R. (1994). A school district's perspective. In T. R. Guskey (Ed.), *High stakes performance assessment: Perspectives on Kentucky's educational reform* (pp. 55-63). Thousand Oaks, CA: Corwin Press.

Polikoff, M. S., Porter, A. C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal*, *48*(4), 965-995.

Popham, W. J. (2006). Needed: A Dose of assessment literacy. *Educational Leadership*, *63*(6), 84-85.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, *48*(1), 4–11.

Popham, W. J. (2011). Transformational assessment in action. Alexandria, VA: Association for Supervision and Curriculum Development.

Popham, W. J. (2018a, June). *Expanding assessment literacy: A pitch to American publishers*. Presentation to the Council of Chief State School Officers 2018 National Conference on Student Assessment, San Diego, CA.

Popham, W. J. (2018b). *Assessment literacy for educators in a hurry.* Alexandria, VA: Association for Supervision and Curriculum Development.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. *Applied Measurement in Education*, *4*(4), 347-362.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Research news and comment: Performance assessments: Political rhetoric and measurement reality. *Educational Researcher*, *21*(2), 22-27.

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, *72*(7), 534-539.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, *77*(3), 238.

Trimble, C. S. (1994). Ensuring educational accountability. In T. R. Guskey (ed.), *High stakes performance assessment: Perspectives on Kentucky's educational reform* (pp. 37-54). Thousand Oaks, CA: Corwin Press.

U.S. Congress (2001). *No Child Left Behind Act of 2001*. Washington, DC: Author.

U.S. Congress (2015). *Every Student Succeeds Act of 2015*. Washington, DC: Author.

Webb, N. L. (2002). Assessment literacy in a standards-based, urban education setting. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

White, R. S. (2018). Who do state policy makers listen to? The case of teacher evaluation. *Phi Delta Kappan*, *99*(8), 13-18.

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*(1), 149-162.

Using Parallel Surveys and Reflective Conversations to Tap Perspectives and Promote Improvement

Giselle O. Martin-Kniep, PhD
President
Learner-Centered Initiatives, Ltd
Garden City, NY

Brett Lane, MA
President,
Institute for Strategic Leading and Learning (INSTLL, LLC)
Catonsville, MD

Abstract

This article focuses on how principals can use parallel surveys that tap multiple perspectives to identify improvement needs, priorities, and the relevance and effectiveness of actions and strategies implemented, and ultimately use this data to support ongoing improvement. We explore the use of survey data that couples principals' self-assessment of their practice with teachers' assessment of those practices, and the value of reports that reconcile the data from these items along with a formal conversation with principals to facilitate reflective use of data. We conclude with recommendations for how to develop school surveys, how to report survey data, and how to facilitate reflective practice and school improvement actions.

Key Words

principal leadership, survey design, data use, principal effectiveness, school effectiveness, school improvement, reflection

Introduction

School surveys are frequently used to understand the perceptions of various stakeholders in a school. Despite the importance of the information they provide, they tend to be underutilized in informing school improvement initiatives. Research concerning the use of survey data by schools and school districts is scarce (Godreau Cimma, 2011) and such research suggests that many leaders are overwhelmed by the amount of available data (Monpas-Huber, 2010).

This article explores how principals can use data from parallel surveys to accelerate improvement. We begin with a discussion of the challenges principals face in using the ever-growing array of data and the important, yet underutilized role of survey data. We then review the technical features of school-based surveys designed to support improvement and two surveys that we have used in such efforts. We close with an example of how a principal used survey data to drive improvement to illustrate key supports that enable effective data use.

Data Use Challenges

Although data are plentiful, principals often lack the knowledge or time to sort through stacks of data from different sources designed to be used in different ways. They also struggle with aligning assessment data with qualitative and quantitative survey data on school organization, culture, and climate.

Principals struggle to formulate data-related questions, limiting their ability to analyze, interpret, and use data effectively. Survey data is often lost in the shuffle of competing "data dives" and an emphasis on summative assessments.

Despite over 30 years of requirements for "data-based improvement planning," school leaders and teachers do not consistently analyze existing data sources within the school, apply such analysis to innovate teaching, curricula, and school performance, and use data to implement and evaluate these innovations (Ingram et al. 2004; Stecker, Fuchs, & Fuchs, 2005). Data—again, summative, formative and benchmark, and survey data—have been mostly used to monitor progress, but outcomes of this monitoring are not consistently applied to improve education (p. 494-495).

In addition to challenges associated with data use, principals and teachers often operate in isolation, with few opportunities to learn from and with each other (Przybylski, 2016).

As a result, little attention is given to the impact of their perspectives of each other's actions on teaching and learning. According to Schildkamp & Kuiper, 2010, data used by school leaders, in most schools, has not led to genuine improvement efforts. Other research (Goldring, et. al. 2015) shows that principals often experience cognitive dissonance when feedback from different data sources (e.g., their self-ratings to those of their teachers) represent conflicting views.

To make better informed decisions, principals would benefit from data that frames the challenges they face and provides different options in resolving them. Deliberate review of qualitative data from surveys and other sources can surface issues that may lead to innovative actions.

Generally, school-level surveys have three, sometimes overlapping, purposes:

(1) **Accountability**: To obtain an objective measure of "school quality" as part of educational accountability, provided by stakeholders (e.g., parents, teachers, students);

(2) **Research**: To measure changes in teacher or student behavior in response to an intervention, or to better understand contextual and curricular aspects of schools; and,

(3) **Improvement**: To inform continuous improvement among district leaders, principals, and teachers.

While each purpose is important, we are focused on the third purpose—improvement— with the goal of examining how principals can use survey data to accelerate improvement efforts. Our experience and research have indicated that there are insufficient supports (e.g. leadership coaching) to assist leaders in making sense of survey data in a manner that leads to improvement and actionable shifts. The questions, then, that we endeavor to answer are:

How can principals use parallel survey data to directly inform improvement-oriented decision and actions?

Specifically:

1. What are the technical features of effective school-level surveys?
2. What systems and supports need to be in place for school leaders to analyze and take actions based on survey information?
3. How can we best facilitate these discussions and actions?
4. What are the types of leadership changes (leadership moves, actions, shifts in behavior) that

leaders may make as a result of reviewing survey data?

Technical Features of Effective School-Level Surveys

A key technical feature of school-level teacher surveys is the use of an evidence-based framework with dimensions that are comprised of individual indicators/items. An evidence-based framework allows users to quickly visualize areas for growth and strength, as well as to track change over time.

While there are differences across surveys, most include categories focused on core instruction, leadership, teacher collaboration, and culture and climate.

Some surveys also include sections on parent and community involvement. Also, most surveys target teachers, students, and sometimes parents and community members. They do not include a separate survey for the principal and leaders although some include administrator responses as part of the teacher survey.

In our work with leaders we have found that it is difficult for principals to actively use teacher survey data without directly referencing, or being able to assess, how their own actions and perceptions mesh with teachers' perceptions.

Background on Survey of Professional Interactions and Organizational Capacity

We initially developed two companion surveys, an Assessment of Professional Interactions and an Assessment of Organizational Capacity, each grounded in the *ARCS Framework for Sustainable School Improvement (see Display 2)*. Each survey was designed from the ground up to provide actionable information for district

and school leaders to inform improvement efforts. The companion teacher surveys were developed in 2009 and have been used in multiple schools (25+ schools) and districts, primarily in New York and Massachusetts.

Our refinement of these surveys highlights three additional, and crucial, technical features of surveys that contributes to the effective use of data.

Specifically, surveys should be capable of: (1) assessing school-level relational networks and the frequency of interactions among individuals; (2) assessing the organizational capacity of the schools; and, (3) allowing principals to compare leader and teacher responses, through parallel principal and teacher items.

The *Assessment of Professional Interactions* gathers information about the frequency and impact of the interactions among school staff, including administrators and district staff, teachers and other professional staff (e.g., guidance counselors, specialists, coaches) focused on teaching and learning. An expanded version of this tool allows for the development of a network map of the connections within a school, identifying key connectors, or "hub individuals" within the school and the density of relationships across teachers, grade-levels, and administrators and coaches. What follows are responses to questions asked by the survey.

Display 1. Assessment of Professional Interactions

	Categories and Scales
What is the **focus** of teachers' professional interactions with each other and with administrators and coaches and how often do you meet around these focus areas?	Professional Discourse Collaboration Instruction Lesson Study Data Use
How **often** do teachers interact with each other, and with administrators and coaches?	With Principal With Assistant Principals With Coaches
Which interactions are deemed to be **most and least useful** and impactful on student learning, and on teachers' professional learning?	With Teachers in same grade With Teachers in other grades With Other Teachers (e.g. SPED or ELL)
What are the **structures** that support the most frequent interactions?	Grade-level Teaming Vertical Teaming Leadership Teams Professional Development (e.g., faculty meeting) Coaching (e.g., job-embedded)

Scales: Daily, Weekly, 2-3 times/month, Less than 1x/month. Very Useful, Somewhat Useful, Not Useful

The *Assessment of Organizational Capacity* measures the school's current capacity to engage in sustainable and effective improvement efforts. This portion of the survey is aligned with the ARCS Framework and produces item-specific results and aggregate ratings for each of the Framework dimensions. Resulting data answers questions related to the ARCS Framework and the extent to which school leadership practices reinforce professional learning and instructional improvement. All items use a 4-point scale (Strongly Agree, Agree, Disagree, Strongly Disagree).

Display 2. The ARCS Framework and Key Questions in each ARCS Dimension

The *ARCS Framework for School Improvement* (Picone-Zocchia and Martin-Kniep, 2009) is based on years of practical experience working with state leaders, district, and schools and evidence-based research, including research on leadership (Leithwood, et. al. 2017) and school improvement (Bryk, et. al. 2015). This framework asserts that the key dimensions of improvement are alignment, representation, culture and sustainability.

Alignment To what extent is curriculum and instruction aligned and how do leadership practice support alignment? How effective is the school leadership in supporting teachers' work and student learning?	**Alignment** examines and questions connections, coherence, focus, direction and sequence among structures, programs, practices and systems. It provides the focus for organizational goal setting, action planning and decision-making, enabling school leaders to strategically connect goals to actions, philosophy and values to practice, and policies to programs and practices.
Representation How aware and involved are staff member in school improvement efforts, including teaming practices, planning, and professional development?	**Representation** examines stakeholder engagement, participation, assumptions, perspectives and constituencies so that they directly inform decision-making and day-to-day actions.
Culture How effective are the structures and practices that ground the development and review of curriculum and use of data to identify and address student needs? To what extent do teachers and leaders share responsibility and hold each other accountable for student learning?	**Culture** focuses on the predominating attitudes, behaviors and beliefs, knowledge and values that characterize schools by examining what people value, what they do and what they produce, focusing on collaborative and reinforcing relationships among participants that promote sharing and learning.
Sustainability How does the school monitor and assess the effectiveness of improvement strategies? How does the school support and mentor teachers and leaders, and plan for staffing turnover?	**Sustainability** frames the school's exploration of its own continuance and meaning beyond the present moment or immediate importance, and shines light on the degree to which it is attending to developing its own expertise, leadership and longevity.

During the first few years of using the survey with schools and sharing data with principals, we noticed that while principals were able to use the survey data to identify areas for improvement, such as grade-levels not working together frequently, or teachers not fully engaged in instructional planning, they did not make connections between their own actions and the experience of teachers, as expressed in the survey data. It was difficult for them to use the data to identify strategic actions or to reconcile survey data with actions that the principal had taken in the past.

To address this dilemma, and in conjunction with our work developing principal leadership evaluation tools and guidance for states, we decided to more explicitly connect teacher perspectives with those of the principal. This shift in our work is reflected in the following theory of action.

Our theory of action proposes that:

If principals had access to elevant data on their leadership moves and practices, and on the relationship between such practices and moves and those of teachers; and, if they had opportunities to reflect on that data and their implications for their systems, structures, processes and practices;

Then, principals would be able to translate the use of such data into meaningful and timely actions to promote improvement for teachers and for the culture of their schools; so that, these actions could lead to improved student outcomes.

Research suggests that principals demonstrating a level of proficiency with the use of a variety of data sources are more adept at designing strategies to address school needs, are more inclined to use data when planning, and are more inclined to initiate goals for school improvement (Przybylski, 2016). When principals are given time, context, and skills to use data, student achievement improves (Schildkamp & Kuiper, 2009[BL2]).

We contend that if principals were more proficient at interpreting data and explore the relationship between their perspectives and those of staff, that all stakeholders in the school would benefit. Gains would be realized in areas such as curriculum development, classroom instruction, remediation and tiered instruction, special needs programming, and professional development opportunities for teachers. Principals would benefit from the data on their practices and the relationship between their practices and teachers' own actions, and from ongoing opportunities to reflect on their practice and on the impact of their actions.

To supplement the teacher survey and support schools' use of survey data, we created a parallel Principal Survey in 2015. The principal survey includes items related to the frequency of the interactions that principals have with others as well as items related to the degree to which they support various aspects of school leadership. The principal survey also assesses the alignment between a principal's individual capacity and overall school vision and the extent to which the school engages in strategic and long-term planning.

Many of the items in the teacher survey and principal survey are parallel to compare teacher responses with leaders' responses to the same questions. Our goal in creating the principal survey was to collect data that could lead to productive conversations with the principal (and leadership) regarding different perspectives around how leadership actions were being understood by teachers, and to explore ways to use this information to craft specific actions.

Table 1

Sample Parallel Teacher and Principal Survey Items (Scale 4, Strongly Agree to 1, Strongly Disagree)

Teacher Items	**Principal Items**
The principal actively engages teachers in promoting the school's instructional focus.	I actively engage teachers in promoting the school's instructional focus.
The principal is knowledgeable about the achievement and progress of every student in the building.	I am able to monitor the achievement and progress of every student in the building.
The principal makes his or her expectations for meeting instructional goals clear to the staff.	I am able to communicate my expectations for meeting instructional goals clearly to the staff.
The principal visits classrooms on a daily basis.	I visit classrooms on a daily basis.
The principal is strongly committed to shared decision making.	I have integrated shared decision making into the leadership of my school.
The instructional feedback that I receive from the principal is useful.	I see evidence that the feedback I give to teachers leads to changes in their practice.
The principal implements processes and structures to ensure quality instructional practices.	I have evidence that the processes and structures I have implemented promote quality instructional practices.
The principal attends to both the learning and social needs of students and staff.	I attend to both the learning and social needs of my staff.
The principal promotes informal and formal leadership opportunities for staff and students.	I actively promote informal and formal leadership opportunities for staff and students.
The principal values reflective practice for him/herself and others.	I cultivate reflective practice in myself and others.
The principal is transparent about the reasoning behind his/her decisions and actions.	I am transparent about the reasoning behind my decisions and actions.
The principal collaborates with staff and other stakeholders around quality teaching and learning.	I collaborate with staff and other stakeholders around quality teaching and learning.

Actions and Supports Needed for Effective Use of Survey Data

The primary purpose of pairing the teacher survey with the principal survey is to stimulate the principal's thinking about the school and his/her work. We provide an illustrative case study to describe how three steps and related facilitation strategies can promote reflection and the development of strategic actions.

Key Step #1. First, it is important to share data reports that link the teachers' and principal's data to see the alignment and gaps in perceptions. These reports include all parallel teacher and survey items in a sortable spreadsheet, so that the principal (and we, as the facilitators of the conversation) can sort high and low scoring items. We also identify items showing differences between principal and teacher responses, highlighting similar and different items. Displaying data according to high and low scoring items and areas of difference is the starting point for the analysis and conversations.

During the first year of our work with West Middle School, survey results were not widely used, despite the inclusion of parallel principal/teacher items. While the survey was provided to leaders with the expectation that the principal would review the data, later conversations with school leaders revealed that the leadership team had reviewed assessment and behavior data to develop their improvement plan for year two, but did not review the survey data.

During year two of the school's improvement effort, the school continued to struggle to implement key improvement initiatives, including having teachers develop high-quality lessons and use grouping strategies in lessons. School leaders also struggled with visiting classrooms to monitor and provide support to teachers. This led to the facilitators'

decision to schedule a formal meeting to review the year two survey data.

Key Step #2. This second key step is to facilitate conversations to unpack the data and promote meaning making. In these conversations, we first review the overall survey data to identify a few key issues, using the following questions:

- How might responses be different if the survey were administered earlier or later in the year?
- Are there different data that would challenge or verify the data from these tools?
- What does the data on teachers' interactions reveal about what teachers' value?
- What does the data on interactions reveal about how the school attends to collaborative work?

We then move on to an analysis of the parallel teacher/principal items, asking questions such as:
- What do the items in which there is a strong alignment in teachers' and principal response reveal about the relationship between teachers and principal?
- What could explain the misalignment between teachers and the principal's responses in the items that reveal misalignment?

In planning for year three, we met with the West Middle School's principal to discuss the results of the survey and how they could be used to inform improvement planning using the preceding questions. We identified *converging* items (e.g., alignment in teachers' and principal responses) and *diverging* (or mis-aligned) items. A sample of these items is provided here.

Converging Items	Teacher Rating	Principal Rating	Difference
The principal visits classrooms on a daily basis.	1.47	2	0.53
Grade-level teams regularly evaluate the effectiveness of instruction through the ongoing analysis of data.	2.00	2	0.00
The school provides every student with appropriate tiered interventions according to needs identified through data.	2.09	2	-0.09

Diverging Items			
The principal values reflective practice for him/herself and others.	2.16	4	1.84
The principal is committed to improving his/her own leadership practices.	1.90	4	2.10
My colleagues are free to bring ideas forward, regardless of their role or formal position.	2.32	4	1.68
Our school implements, monitors and evaluates the impact of vertically aligned instructional strategies.	1.45	3	1.55
The school commits to targeted goals and priorities and sticks with them over a long period of time.	1.55	3	1.45
The school encourages adults to work in groups and teams to learn from each other.	2.45	4	1.55

Our unpacking of these items (and others) provided clarity on areas for improvement and highlighted potential explanations for why improvement efforts had not been as successful as anticipated. For instance, the relatively low and aligned perceptions of teachers regarding the frequency of principal visits, grade-level analysis of data, and provision of tiered interventions, clarified the need to improve these structures and practices.

While this information was not necessarily new, since it reflected challenges that were part of the school's improvement plan, our in-person analysis of diverging items began to surface explanations as to why improvement efforts were not successful.

Key Step #3: Third, after unpacking the areas of convergence and difference between teacher and principal responses, we focus on a few of the diverging items. Instead of dwelling on explanations as to *why* there are differences, we *ask the principal to consider what actions he/she could take to change teachers' perceptions*. We highlight this as an important "facilitation move" that redirects the conversation from focusing on why teachers may have different perceptions, which can lead to potentially <u>defensive or non-productive</u> responses (e.g., teachers don't know what I do; teachers don't understand the full picture) to a more proactive conversation, focusing on what the principal can do to change perceptions.

In West Middle School, the principal found many of diverging items to be unsettling, noting that her teachers did not acknowledge her efforts to be reflective, encourage collaboration among colleagues, and set grade-level and school goals. In our conversation, the principal noted that she had worked hard to put into place teaming structures (for collaboration)

Specifically, the principal rated herself highly in terms of valuing reflective practice, setting targeted goals, and encouraging collaboration and sharing of ideas among staff. However, staff did not perceive the principal as doing so.

We assert that such "disconnect" between the values and perceptions of leaders and staff may be common among low performing schools that are working hard to change and improve, and that addressing this disconnect is essential to sustained and effective improvement efforts. Rather than framing improvement strategies as simply "effective or ineffective," the successful implementation of improvement strategies rests upon reconciling different perceptions, values, and understandings among stakeholders. and that many of her actions were directed towards school improvement goals.

As we discussed actions the principal could take to change teachers' perceptions, we uncovered that teachers may not fully understand how the principal's actions (e.g., directives, allocation of staff, changes in teaming structures) related to building collaboration and trust, or to overall school goals. This conversation contributed to specific principal actions to clarify the connections between her actions, school teaming structures, and roles and responsibilities of coaches, department heads, and teachers.

The principal and her leadership team took deliberate and strategic actions in year three. For instance, she explicitly communicated why certain actions were taken, such as asking teachers to submit lessons to a shared google drive and to incorporate 20 minutes of group work into each instructional period. In the past, such actions had been construed by teachers as a top-down attempt to "monitor" or control teacher practice. Taking

the time to explain how these actions aligned with the schoolwide goal of developing instructional coherence provided credible justification for changes in teacher practice. This message was reinforced by coaches and department heads. Similarly, the principal developed a personal schedule of informal classroom visitations and informed teachers that she was doing so to hold herself accountable and to ensure that teachers received feedback.

Another example of proactive communication was the principal's clarification to the staff of the roles of Department Heads with respect to content teams, which connected the work of instructional coaches (responsible for grade-level common planning) with the responsibilities of department heads (to develop and monitor strong content).

In year three, we noted shifts in principal actions and teacher responses, a direct result of our analysis of the survey and the principal's diligence in cultivating shared leadership and communication. Examples of shifts include:

- The development, sharing, and use of an informal walkthrough tool with department heads and teachers that included specific "look-fors" related to student engagement, higher order thinking, and student grouping—all key aspects of the school's instructional model. The walkthrough tool was subsequently used by the principal and department heads to visit all classrooms on a weekly basis, to proactively identify teachers needing additional instructional support and to inform schoolwide professional development.
- The inclusion of a similar set "look-fors" on a shared lesson planning template that made explicit expectations for lessons and provided a way for coaches and department heads to review lesson plans and provide feedback to teachers.
- By setting clear expectations for grade-level teams and PLCs as time for teachers to develop common lessons on shared practices, with support from coaches and department heads. Teachers now have a growing ownership of shared instructional practices and a willingness to develop common lessons and share best practices (and challenges) related to student groupings and formative assessments.

Conclusion

Quantitative and qualitative data on teachers' practices and student outcomes are insufficient in terms of informing the questions, decisions and actions that principals need to make to improve their schools.

If principals truly want to build a culture of trust that leads to school improvement, they need to compare their own perceptions with data on how teachers perceive them. The proactive use of parallel survey data could greatly enhance and complement principals' reliance on test data, teacher observation and other quantitative data sources, and could lead to better and deeper analysis of existing data sources.

Given the overwhelming amount of data principals have access to, and the fact that data sources are seldom integrated into accessible reports, principals could benefit greatly from formal opportunities to explicitly assess the data they have; ideally facilitated by external providers.

There is great value in conversations that enable principals to identify and reconcile differing perspectives to test their own

assumptions and consider their school system through the lenses of others with an open mind. Without the conversation, principals may not be able to take needed actions.

Further work is needed to determine how to best engage leaders in the reflective analysis and use of data to make this process cost effective. It may be useful to incorporate the analysis and use of survey data into school improvement efforts, especially those that are externally facilitated, or are mandated to low-performing schools.

Author Biographies

Giselle Martin-Kniep is the founder and president of Learner-Centered Initiatives, Ltd. She has a background in program evaluation, organizational change, and political science and has several graduate degrees from Stanford University. Over the past 25 years, she has worked with national and international schools and districts in the areas of strategic planning, curriculum and assessment, educational policy, neuro leadership, and systems thinking. Most recent interests center on determining best leverage points for sustainable organizational improvement, and more specifically around developing and aligning outcomes and measures. E-mail: gisellemk@lciltd.org

Brett Lane is the president of INSTLL, an education research and policy organization focused on promoting innovative and meaningful change in public education. He has an MA in political science and 25 years of direct work with state, district, and school leaders. He has worked for two regional education laboratories and, over the past 10 years, as a partner with states (Massachusetts and Michigan) and multiple school districts. Research and consultative work have focused on efforts to scale up improvement efforts across schools and districts through policy and statewide systems that provide the conditions needed for effective and sustainable improvement. He is currently researching and working on implementing district improvement and turnaround efforts through networked improvement communities. E-mail: brett.lane@instll.com

References

Bryk, A.S., Gomez, L.M., Grunow, A., & LeMahieu, P. G. (2015) *Learning to Improve: How America's Schools Can Get Better at Getting Better*. Boston, MA: Harvard Education Press.

Godreau Cimma, K. L. (2011). A middle school principal's and teachers' perceptions of leadership practices in data-driven decision making. (Ed.D., University of Hartford). *ProQuest Dissertations and Theses.(*851186662).

Golding E., Mayrogordato, M., Taylor Haynes, K. (2015) Multisource Principal Evaluation Data: Principals' Orientations and Reactions to Teacher Feedback Regarding Their Leadership Effectiveness. *Education Administration Quarterly*. Volume: 51 issue: 4, page(s): 572-599

Ingram, D., Louis, S. K., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: barriers to the use of data to improve practice. *Teachers College Record*, 106(6), 1258–1287

Leithwood, K., Sun, J., & Pollock, K. (Eds.). (2017). *How school leaders contribute to student success: the four paths framework*. Cham: Springer International Publishing.

Monpas-Huber, J. B. (2010). Explaining teachers' instructional use of state assessment data: A multilevel study of high school teachers in Washington state. *Journal of School Leadership*, 20(2), 208-237.

Picone-Zocchia, J., and Martin-Kniep, G. (2009) *ARCS Framework for Sustainable School Improvement.* NY: Communities for Learning. http://www.lciltd.org/resources/the-arcs-framework

Przbylski, R.M. (2016) I have the information, now what? A study discovering the factors that make up data literacy/competence. Doctoral dissertation. The University of Alabama, *ProQuest Dissertations Publishing*. 10127855.

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, 26, 482-496.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: review of research. *Psychology in the Schools,* 42(8), 795–819

Sun, J.- P., Przybylski, R. M., & Johnson, B. (2014). How to promote teachers' use of student data. Manuscript submitted to *Studies in Educational Evaluation*.

---

[1] We are focusing on technical features that relate to the content and usability of surveys, rather than issues related to the validity and reliability of survey items.

All Value-Added Models (VAMs) Are Wrong, but Sometimes They May Be Useful

Audrey Amrein-Beardsley, PhD
Professor
Mary Lou Fulton Teachers College
Arizona State University
Tempe, AZ

Edward Sloat, EdD
Faculty Associate
Mary Lou Fulton Teachers College
Arizona State University
Tempe, AZ

Jessica Holloway, PhD
Australian Research Council DECRA Fellow
Research for Educational Impact
Deakin University
Melbourne, Australia

Abstract

In this study, researchers compared the concordance of teacher-level effectiveness ratings derived via six common generalized value-added model (VAM) approaches including a (1) student growth percentile (SGP) model, (2) value-added linear regression model (VALRM), (3) value-added hierarchical linear model (VAHLM), (4) simple difference (gain) score model, (5) rubric-based performance level (growth) model, and (6) simple criterion (percent passing) model. The study sample included fourth to sixth grade teachers employed in a large, suburban school district who taught the same sets of students, at the same time, and for whom a consistent set of achievement measures and background variables were available. Findings indicate that ratings significantly and substantively differed depending upon the methodological approach used. Findings, accordingly, bring into question the validity of the inferences based on such estimates, especially when high-stakes decisions are made about teachers as based on estimates measured via different, albeit popular methods across different school districts and states.

Key Words

teacher accountability, teacher evaluation, teacher effectiveness, teacher assessment, validity/reliability, value-added models (VAMs)

Statistical Philosophy

In 1976, British statistician George Box remarked: "Essentially, all [statistical] models are wrong, but some are useful." What Box argued was that statistical models have to be understood for what they can do, which is *estimate* that which a scientist is attempting to measure. Indeed, statistical models never yield *true* measures of anything.

This is particularly important in education, as we are currently facing a "data explosion" (SAS, n.d.) where statistical models are being used to measure just about any and all conceivable matters, including teacher performance. Unfortunately, this new data milieu has brought about dangerous applications of data and statistics.

We say dangerous because complex matters are too often drastically oversimplified for measurement's sake, and when we oversimplify complex matters, we run the very real risk of making erroneous inferences that overlook important considerations, conditions, and circumstance that may lead to consequential decisions that are incorrect.

At the same time, however, useful insights can be gleaned from statistical models, even if they do not offer *true* representations of that which they are used to model. It goes without saying that probability-based prediction models are essential tools in other disciplines including business, medicine, manufacturing, and meteorology, just to name a few. Yet, no matter how historically accurate statistical models may be, all statistical predictions are imperfect.

Notwithstanding, as long as those who consume and interpret model output understand the imperfections at play, and they do not exaggerate the degree to which statistical models might provide useful information, then models can sometimes offer valuable insights about social phenomena. If we accept that statistical perfection is impossible, then we can begin to consider, perhaps and according to Clear (2018), "whether something can be applied to everyday life in a useful manner." A recent study we conducted directly dealt with this provocation, and what we found was that statistical models that measure teacher effectiveness are, in the words of Box, essentially wrong, but sometimes useful when critically consumed or used.

In our study, we explored the use of different value-added models (VAMs, see more forthcoming) to evaluate teachers' measurable impacts on their students' test scores (see also Sloat, Amrein-Beardsley, & Holloway, 2018). Findings should be of great interest to school leaders throughout the U.S. who continue to struggle with what they can and cannot do with these potentially problematic statistical data.

Ideally, findings from this study should help school leaders better understand how VAMs can be used for making important decisions about their schools and teachers, as well as where school leaders might draw the line about the consequences they attach to VAM output. The key takeaway for school leaders is to *not* place high value on these statistical measures, but rather devalue them as much as possible.

Past and Current Circumstances

Since the federal government's Race to the Top (RttT) Act of 2011, and the No Child Left Behind (NCLB, 2001) waivers that excused states from penalties associated with their failures to meet NCLB's 100% student proficiency goals by 2014, most states (and districts) have developed and used teacher evaluation systems that rely in large part on student test scores to "objectively" measure and evaluate teacher effectiveness. While the

federal passage of the *Every Student Succeeds Act* (ESSA, 2016) has since helped curb such educational accountability and reform efforts, particularly at the teacher level, ESSA continues to encourage states to hold teachers accountable for that which statistically matters, including their students' test scores.

Consequently, teacher performance is still being calculated using complex statistical modeling approaches and practices, primarily via VAMs. VAMs, in the simplest of terms, classify teachers' effectiveness levels according to their statistically measurable and purportedly causal impacts on their students' standardized test scores over time.

Ideally, VAMs help to identify teachers whose students outperform their projected levels of growth as effective and teachers whose students fall short as ineffective. In reality, however, VAMs often do not work as intended, raising questions about whether VAM-based data can be used as objective measures for teacher evaluation purposes (see, for example, Amrein-Beardsley, 2014).

The Study

To address this concern, we conducted a study comparing the concordance, or rather the agreement of VAM scores across six different VAMs. More specifically, researchers compared the concordance of teacher-level effectiveness ratings derived via six common generalized VAM approaches including a: (1) student growth percentile (SGP) model, (2) value-added linear regression model (VALRM), (3) value-added hierarchical linear model (VAHLM), (4) simple difference (gain) score model, (5) rubric-based performance level (growth) model, and (6) simple criterion (percent passing) model.

For each approach, researchers used the distribution of teacher-level estimates by

subject area to rank teacher effects and then assign them effectiveness ratings. Thereafter, researchers statistically evaluated the level of agreement between and among ratings to examine concordance, with concordance statistically approximated by the extent to which similar results and conclusions were drawn, via these independent methods with common purpose. The overall intent was to examine what impact the choice of the methods implemented, as locally defined, would have on the inferential and potentially consequential judgments of effectiveness made.

The primary research question researchers investigated was to what extent teacher-level ratings significantly or substantively differed depending upon the methodological approaches used, with concordance yielding evidence of criterion-related evidence of validity and a lack of concordance the inverse, while also bringing into question the validity of the inferences based on such estimates especially when high-stakes decisions are to be attached to such estimates.

Researchers defined concurrent concordance via statistical approximations of the extent to which similar results for the same teachers at the same time were drawn via independent, common, and more generalized VAMs (for more detailed, technical information, please see Sloat et al., 2018).

Findings

We found that teachers' ratings significantly differed 18%-59% of the time depending on the VAM used. What this means is that, even when using the same data, from the same tests, for the same students, and for the same teachers, different VAMs produced very different teacher effectiveness scores. That is, a single teacher could be classified differently depending on which VAM was used. This is

critical when considering the *efficacy* and *ethics* of whether VAMs should be used for teacher evaluation purposes. Likewise, not only does the seemingly simple choice of which VAM a school district might use become nettlesome, so too does the question about whether we can trust really any VAM for high-stakes purposes. While the differences in VAM results might not matter as much if used for low-stakes purposes (e.g., making professional development decisions for certain sets of teachers), they certainly matter a great deal if used for matters like teacher tenure decisions, merit pay, teacher probation and termination.

Findings from this study, consequently, bring into serious question the *validity* or *truthfulness* of the inferences based on VAM estimates, especially when high-stakes decisions are made about teachers.

Furthermore, because many school districts, especially small districts, districts located in certain urban or rural areas, American Indian districts, and the like, do not have comparable access to the in-house expertise (e.g., data analytics, statistical methods) or resources (e.g., hard/software, data management systems) necessary to support even a run-of-the-mill statistical model of teachers' effects (i.e., a VAM), different results might also be related to financial and human resources more than teachers' *true* effects.

That some districts will rely upon simplistic metrics of teacher effects is also deeply problematic as how a teacher is evaluated greatly depends on the approach the district chooses. This choice is heavily constrained by the district's technical capabilities, as well as human and technical resources, threatening the core validity of any inference derived from the chosen method. If different methods yield different outcomes,

then the truthfulness of the inferences and any related decisions to be made are warped.

Implications for School Leaders

The fact that different VAMs produce different results is indeed alarming, but there are some important caveats to consider. With ESSA (2016) now legislating that districts can determine which VAM they might adopt, teachers' classifications will depend upon whichever model their district has chosen to implement, making this whole statistical modeling enterprise arbitrary across varying contexts.

However, we are not suggesting that one statistical model be adopted for all districts for purposes of consistency, for there is really no professional consensus that any particular VAM is better or more accurate than any other (although VAM proprietors would likely disagree). Recall that all VAMs are reliant upon statistical models that only *estimate*, as best they can, that which is an accurate representation of *truth*.

Consequently, that where a teacher teaches, and what value-added method is used in that district, might matter more than his/her *actual* effectiveness is highly problematic. This in and of itself puts at risk the validity of such teacher-level accountability outcomes. This also places school leaders in a challenging position, as they must be critically aware of not only the different types of statistical approaches from which to choose, but also of how they might consume, interpret, and act upon the outputs drawn from such models.

Truth be told, all school administrators should be aware that all VAMs yield quite varied estimates of teacher effectiveness, none of which are ever actually *true*. Likewise, they cannot afford to be ambivalent about how

VAM output might be used within their schools, especially if high-stakes consequences are at stake.

When VAMs Are Wrong

Across the U.S. are a series of ongoing or recently completed lawsuits where teacher plaintiffs are contesting how they are being evaluated by VAMs. For all of these cases, teacher plaintiffs are targeting the value-added indicators being used, as alleged, erroneously and inappropriately against them. More specifically, plaintiffs are arguing that multiple VAMs (like those analyzed in this study), are grossly imperfect, arbitrary, capricious, irrational, and unfair (see, for example, Paige, Amrein-Beardsley, & Collin, in press).

Related, plaintiffs are arguing that the preponderant use of VAM-based indicators is more egregious when high-stakes decisions are attached to value-added output. As the stakes increase, the more egregious the actions attached to VAM output. The high-stakes decisions at issue across these specific cases include but are not limited to teachers' permanent files being flagged with their VAM-based effectiveness categories (e.g., "highly effective," "effective," "ineffective," "highly ineffective") that has prevented teachers from moving teaching positions across districts; the awarding or revocation of teacher licenses or tenure; salary increases, decreases, or merit pay; and teacher probation or termination.

Most notable across suits, though, are a few cases that quite literally make the case we are making here, about when VAM use is simply wrong, now also as per the courts. In Houston in 2011, 221 teachers were terminated as based predominantly on their VAM scores. A U.S. District Court ultimately ruled in favor of teacher plaintiffs in this case given they had legitimate claims regarding how the VAM

being used by the district violated their Fourteenth Amendment due process protections, more expressly given the district's VAM did not permit district teachers to ensure their VAM scores were accurate. The district got rid of their VAM.

In New Mexico, despite a widespread understanding that teachers' VAM-based data were to be held "harmless" until teachers' VAM data could be studied, vetted, and validated, the state flagged teachers' permanent files, as mentioned prior, with teachers' VAM-based effectiveness categories. This ultimately prevented some teachers from moving teaching positions across districts within the state. This landed the state and its statewide VAM in court.

A State Court judge ultimately granted a preliminary injunction to prevent the state or any district within the state from making any consequential decisions about New Mexico teachers until the state could evidence that such consequences as attached to the state's VAM were warranted, non-arbitrary, legally defensible, and "uniform and objective" as per state constitutional requirements. No such evidence has yet been presented to warrant the attachment of high-stakes decisions to teachers' VAM scores, leaving the state at a standstill in terms of its VAM-based teacher evaluation system since 2015.

In New York, the State Supreme Court viewed the consequences attached to its VAM differently, positioning an "ineffective" teacher effectiveness tag as consequential in terms of public shame and loss of reputation in the professional community. The Court ultimately ruled that the state's VAM-based teacher evaluation system was "arbitrary and capricious," defined as actions "taken without sound basis in reason or regard to the facts"

(State of New York Supreme Court, 2016, p. 11).

These cases demonstrate how the U.S. judicial system has thus far interpreted VAMs and VAM use as legally defensible in practice, when high-stakes consequences have been attached to VAM output. While not all cases have been ruled in favor of teacher plaintiffs (e.g., in Tennessee a U.S. District Court dismissed a case given the state's use of its VAM was "rationally related to a legitimate government interest"), the majority have. In fact, a majority of court rulings have reversed states' and districts' high-stakes use of VAMs in that no defendant has been able to produce evidence demonstrating their VAM can produce outputs that warrant high-stakes use.

Accordingly, understanding the value of some of the on-the-ground consequences of VAM use is germane to our collective understandings about these statistical models, in some ways regardless of the different estimates that different models yield. This is also important, again, as many states and districts continue to employ VAM-based evaluation systems despite the serious measurement and pragmatic issues at play, especially when consequential decisions are also at play.

When VAMs May Be Useful

As we noted prior, different VAM-based results (as evidenced in our study) might not matter as much if VAM-based output are used for low-stakes purposes, such as making professional development decisions. Hence, we also want to emphasize, particularly for school leaders, that VAMs may still be useful despite their (oft-gross) statistical shortcomings.

Susan Moore Johnson, professor of education at Harvard University, and some of her colleagues recently published an important article regarding how teacher evaluation systems might actually be useful within school districts. Explained in their article titled *Investing in Development: Six High-Performing, High-Poverty Schools Implement the Massachusetts Teacher Evaluation Policy,* Reinhorn, Moore Johnson, and Simon (2017) "studied how six high-performing, high-poverty [and traditional, charter, under state supervision] schools in one large Massachusetts city implemented the state's new teacher evaluation policy" (p. 383).

They aimed to learn how these "successful" schools, with "success" defined by the state's accountability ranking per school along with each school's "public reputation," approached the state's teacher evaluation system and its system components. They also looked at how the educators in these schools used their evaluation data to promote more opportunities for development.

They found that across the six successful schools that they studied, school administrators "responded to the state evaluation policy in remarkably similar ways, *giving priority to the goal of development over accountability* [emphasis added]" (p. 385). In addition, most school administrators of said successful schools went above and beyond to provide teachers with more frequent observations, feedback, and teacher evaluation supports than any state or district policy required. "Teachers widely corroborated their principal's reports that evaluation in their school was meant to improve their performance and they strongly endorsed that priority" (p. 385).

Overall, the researchers concluded that "an evaluation policy focusing on teachers' development can be effectively implemented in ways that serve the interests of schools, students, and teachers" (p. 402). This is

especially true when (1) evaluation efforts are "well grounded in the observations, feedback, and support of a formative evaluation process," which could include the use of VAM-based data for formative versus summative (e.g., outcome-or accountability-based) purposes; (2) when school administrators focus on "capacity building;" and (3) when states and districts do not take Draconian (i.e., strict or drastic) but judicious and admonitory approaches to teacher evaluation systems and the data they derive.

Developmental and formatively-focused teacher evaluation systems work, they conclude perhaps most importantly, when schools are led by highly effective school leaders. This "is probably the most important thing district officials can do to ensure that teacher evaluation will be a constructive, productive process" (p. 403).

Findings from this study matter in that they offer evidence that teacher evaluation works *if* used for developmental and formative purposes, perhaps in lieu of summative and despite high-stakes purposes and demands.

Current evidence also suggests that post-ESSA (2016) nearly all states are moving in this direction (Close, Amrein-Beardsley, & Collins, 2018). States' new teacher evaluation plans make note of providing data to teachers as a means of supporting professional development and improvement, essentially shifting the purpose of the evaluation system away from summative and toward formative use.

Final Remarks

Despite ESSA, many teacher evaluation systems still include VAMs. What is important is that as long as the output are consumed and interpreted critically in terms of VAM's strengths and weaknesses, and they are used for formative versus summative or punitive purposes by school leadrs, then teacher evaluation can work. Of key priority should be that VAM-based and other teacher evaluation data are understood and used relative to their potentials and limitations, and, most importantly, for developmental and formative purposes only.

While the type of statistical modeling used within VAMs is still often referred to as the most sophisticated means for measuring the amount of influence an individual teacher has on his/her students' achievement test scores, as Lingard (2011) argued "The knowledge we produce is … partial, positioned and provisional with limitations when applied as an evidence base" (p. 358). This caveat is important to keep in mind when considering the implications of VAM-based use, especially by school administrators at the district level.

While the data produced by VAMs might be statistically sophisticated, contextual factors will *always* affect how VAMs play out in practice; hence, school administrators and teachers should be armed with as much knowledge as possible about when, why, and how VAMs should be used.

Author Biographies

Audrey Amrein-Beardsley is a professor in the Mary Lou Fulton Teachers College at Arizona State University. Her research focuses on the use of value-added models (VAMs) in and across states before and since the passage of the *Every Student Succeeds Act* (ESSA). More specifically, she is conducting validation studies on multiple system components, as well as serving as an expert witness in many legal cases surrounding the (mis)use of VAM-based output. Email: audrey.beardsley@asu.edu

Edward Sloat is currently a faculty associate in the Mary Lou Fulton Teachers College at Arizona State University. He currently teaches advanced statistical courses and conducts research on contemporary teacher evaluation systems. Formerly, he served in multiple district-level leadership roles overseeing districts' research, evaluation, and accountability divisions. Email: esloat@asu.edu

Jessica Holloway is currently an Australian Research Council DECRA Fellow within the strategic research centre (Research for Educational Impact—(REDI) at Deakin University in Melbourne, Australia. Via her current DECRA project, "The Role of Teacher Expertise, Authority and Professionalism in Education," she investigates the role of education in modern democratic societies, with a particular focus on teachers and teacher expertise. In general, her research draws on post-structural theory to understand contemporary modes of accountability and its production of new teacher and leader subjectivities. Email: jessica.holloway@deakin.edu.au

References

Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability.* New York, NY: Routledge.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association, 71*(356), 791-799. Retrieved from http://mkweb.bcgsc.ca/pointsofsignificance/img/Boxonmaths.pdf

Clear, J. (2018). All models are wrong, some are useful. *JamesClear.com.* Retrieved from https://jamesclear.com/all-models-are-wrong

Close, K., Amrein-Beardsley, A., & Collins, C. (2018). *State-level assessments and teacher evaluation systems after the passage of the Every Student Succeeds Act: Some steps in the right direction.* Boulder, CO: Nation Education Policy Center (NEPC). Retrieved from http://nepc.colorado.edu/publication/state-assessment

Every Student Succeeds Act (ESSA) of 2015, Pub. L. No. 114-95, § 129 Stat. 1802. (2016). Retrieved from https://www.gpo.gov/fdsys/pkg/BILLS-114s1177enr/pdf/BILLS-114s1177enr.pdf

Lingard, B. (2011). Policy as numbers: Ac/counting for educational research. *The Australian Educational Researcher*, *38*(4), 355-382.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115 Stat. 1425. (2002). Retrieved from http://www.ed.gov/legislation/ESEA02/

Paige, M. A., Amrein-Beardsley, A., & *Close, K. (in press). Tennessee's national impact on teacher evaluation law & policy: An assessment of value-added model litigation. *Tennessee Journal of Law & Policy.*

Reinhorn, S. K., Moore Johnson, S., & Simon, N. S. (2017). *Educational Evaluation and Policy Analysis, 39*(3), 383–406. doi:10.3102/0162373717690605 Retrieved from https://projectngt.gse.harvard.edu/files/gse-projectngt/files/eval_041916_unblinded.pdf

Sloat, E., Amrein-Beardsley, A., Holloway, J. (2018). Different teacher-level effectiveness estimates, different results: Inter-model concordance across six generalized value-added models (VAMs). *Educational Assessment, Evaluation and Accountability.* doi:10.1007/s11092-018-9283-7

State of New York Supreme Court. (2016). *Decision, order and judgment.* Roger D. McDonough, Justice. Retrieved from http://vamboozled.com/wp-content/uploads/2016/05/ruling.pdf

Mission and Scope, Copyright, Privacy, Ethics, Upcoming Themes, Author Guidelines, Submissions, Publication Rates & Publication Timeline

The *AASA Journal of Scholarship and Practice* is a refereed, blind-reviewed, quarterly journal with a focus on research and evidence-based practice that advance the profession of education administration.

Mission and Scope

The **mission** of the Journal is to provide peer-reviewed, user-friendly, and methodologically sound research that practicing school and district administrations can use to take action and that higher education faculty can use to prepare future school and district administrators. The Journal publishes accepted manuscripts in the following categories: (1) Evidence-based Practice, (2) Original Research, (3) Research-informed Commentary, and (4) Book Reviews.

The **scope** for submissions focuses on the intersection of five factors of school and district administration: (a) administrators, (b) teachers, (c) students, (d) subject matter, and (e) settings. The Journal encourages submissions that focus on the intersection of factors a-e. The Journal discourages submissions that focus only on personal reflections and opinions.

Copyright

   Authors please note: By submitting a manuscript the author/s acknowledge that the submitted manuscript is not under review by any other publisher or society, and the manuscript represents original work completed by the authors and not previously published as per professional ethics based on APA guidelines, most recent edition. By submitting a manuscript, authors agree to transfer without charge the following rights to AASA, its publications, and especially the *AASA Journal of Scholarship and Practice* upon acceptance of the manuscript. The *AASA Journal of Scholarship and Practice* is indexed by several services and is also a member of the Directory of Open Access Journals. This means there is worldwide access to all content. Authors must agree to first worldwide serial publication rights and the right for the *AASA Journal of Scholarship and Practice* and AASA to grant permissions for use of works as the editors judge appropriate for the redistribution, repackaging, and/or marketing of all works and any metadata associated with the works in professional indexing and reference services. Any revenues received by AASA and the *AASA Journal of Scholarship and Practice* from redistribution are used to support the continued marketing, publication, and distribution of articles.

Privacy
The names and e-mail addresses entered in this journal site will be used exclusively for the stated purposes of this journal and will not be made available for any other purpose or to any other party. Please note that the journal is available, via the Internet at no cost, to audiences around the world. Authors' names and e-mail addresses are posted for each article. Authors who agree to have their manuscripts published in the *AASA Journal of Scholarship and Practice* agree to have their names and e-mail addresses posted on their articles for public viewing.

Ethics
The *AASA Journal of Scholarship and Practice* uses a double-blind peer-review process to maintain scientific integrity of its published materials. Peer-reviewed articles are one hallmark of the scientific method and the *AASA Journal of Scholarship and Practice* believes in the importance of maintaining the integrity of the scientific process in order to bring high quality literature to the education leadership community. We expect our authors to follow the same ethical guidelines. We refer readers to the latest edition of the APA Style Guide to review the ethical expectations for publication in a scholarly journal.

Upcoming Themes and Topics of Interest
Below are themes and areas of interest for publication cycles.
1. Governance, Funding, and Control of Public Education
2. Federal Education Policy and the Future of Public Education
3. Federal, State, and Local Governmental Relationships
4. Teacher Quality (e.g. hiring, assessment, evaluation, development, and compensation of teachers)
5. School Administrator Quality (e.g. hiring, preparation, assessment, evaluation, development, and compensation of principals and other school administrators)
6. Data and Information Systems (for both summative and formative evaluative purposes)
7. Charter Schools and Other Alternatives to Public Schools
8. Turning Around Low-Performing Schools and Districts
9. Large Scale Assessment Policy and Programs
10. Curriculum and Instruction
11. School Reform Policies
12. Financial Issues

Submissions
Length of manuscripts should be as follows: Research and evidence-based practice articles between **2,800** and **4,800** words; commentaries between **1,600** and **3,800** words; book and media reviews between **400** and **800** words. Articles, commentaries, book and media reviews, citations and references are to follow the *Publication Manual of the American Psychological Association*, latest edition. Permission to use previously copyrighted materials is the responsibility of the author, not the *AASA Journal of Scholarship and Practice*.

Cover page checklist:
1. title of the article:
   identify if the submission is original research, evidence-based practice, commentary, or book review
2. contributor name(s)
3. terminal degree
4. academic rank
5. department
6. college or university
7. city, state
8. telephone and fax numbers
9. e-mail address
10. 120-word abstract that conforms to APA style
11. six to eight key words that reflect the essence of the submission
12. 40-word biographical sketch

Please do not submit page numbers in headers or footers. Rather than use footnotes, it is preferred authors embed footnote content in the body of the article. Articles are to be submitted to the editor by e-mail as an electronic attachment in Microsoft Word, Times New Roman, 12 Font. **New**: the editors have also determined to follow APA guidelines by adding two spaces after a period.

Acceptance Rates

The *AASA Journal of Scholarship and Practice* maintains of record of acceptance rates for each of the quarterly issues published annually. The percentage of acceptance rates since 2010 is as follows:

2012: 22%	2016: 19%
2013: 15%	2017: 20%
2014: 20%	2018: 19%
2015: 22%	2019: 19%

Book Review Guidelines

Book review guidelines should adhere to the author guidelines as found above. The format of the book review is to include the following:
- Full title of book
- Author
- Publisher, city, state, year, # of pages, price
- Name and affiliation of reviewer
- Contact information for reviewer: address, city, state, zip code, e-mail address, telephone and fax
- Reviewer biography
- Date of submission

Publication Timeline

Issue	Deadline to Submit Articles	Notification to Authors of Editorial Review Board Decisions	To AASA for Formatting and Editing	Issue Available on AASA website
Spring	October 1	January 1	February 15	April 1
Summer	February 1	April 1	May 15	July1
Fall	May 1	July 1	August 15	October 1
Winter	August 1	October 1	November 15	January 15

Additional Information

Contributors will be notified of editorial board decisions within eight weeks of receipt of papers at the editorial office. Articles to be returned must be accompanied by a postage-paid, self-addressed envelope.

The *AASA Journal of Scholarship and Practice* reserves the right to make minor editorial changes without seeking approval from contributors.

Materials published in the *AASA Journal of Scholarship and Practice* do not constitute endorsement of the content or conclusions presented.

The Journal is listed in Cabell's Directory of Publishing Opportunities. Articles are also archived in the ERIC collection. The Journal is available on the Internet and considered an open access document.

Editor

Kenneth Mitchell, EdD
AASA Journal of Scholarship and Practice
Submit articles electronically: kenneth.mitchell@mville.edu

To contact by postal mail:
Dr. Ken Mitchell
Associate Professor
School of Education
Manhattanville College
2900 Purchase Street
Purchase, NY 10577

AASA Resources

- ✓ **Join AASA** and discover a number of resources reserved exclusively for members. See Member Benefits at www.aasa.org/welcome/index.aspx. For questions on membership contact Chris Daw, cdaw@aasa.org. For questions on governance and/or state relations contact Noelle Ellerson Ng at nellerson@aasa.org.

- ✓ For **Resources on COVID-19**, see https://bit.ly/2xyrcQV

- ✓ For information on AASA's **Community Emergency Management, School Safety and Crisis Planning** go to https://bit.ly/2xyrcQV

- ✓ **Resources for School Administrators** may be seen at www.aasa.org/welcome/resources.aspx

- ✓ The **AASA's Leadership Network** drives superintendent success, innovation and growth, shaping the future of public education while preparing students for what's next. Passionate and committed, the Network connects educational leaders to the professional learning, leadership development, relationships and partnerships through a variety of ongoing academies, cohorts, consortiums, and programs needed to ensure a long career of impact**.** For additional information on leadership opportunities and options visit www.aasa.org/LeadershipNetwork or contact Mort Sherman at msherman@aasa.org or Valerie Truesdale at vtruesdale@aasa.org.

- ✓ For information on AASA's **Future-Focused Schools Collaborative** refer to www.aasa.org/AASACollaborative.aspx

- ✓ For AASA's National Conference on Education (2/21) **Call for Proposals** due May 29, 2020 go to https://www.aasa.org/2021cfp.aspx

- ✓ Learn about AASA's **books program** where new titles and special discounts are available to AASA members. The AASA publications catalog may be downloaded at www.aasa.org/books.aspx.

- ✓ **Virtual Meetings** for EdLeaders may be found at www.aasa.org/AASA-LeadershipNetwork-webinars.aspx

Upcoming AASA Events

AASA 2021 National Conference on Education, Feb. 15-17, 2021 in New Orleans, La.