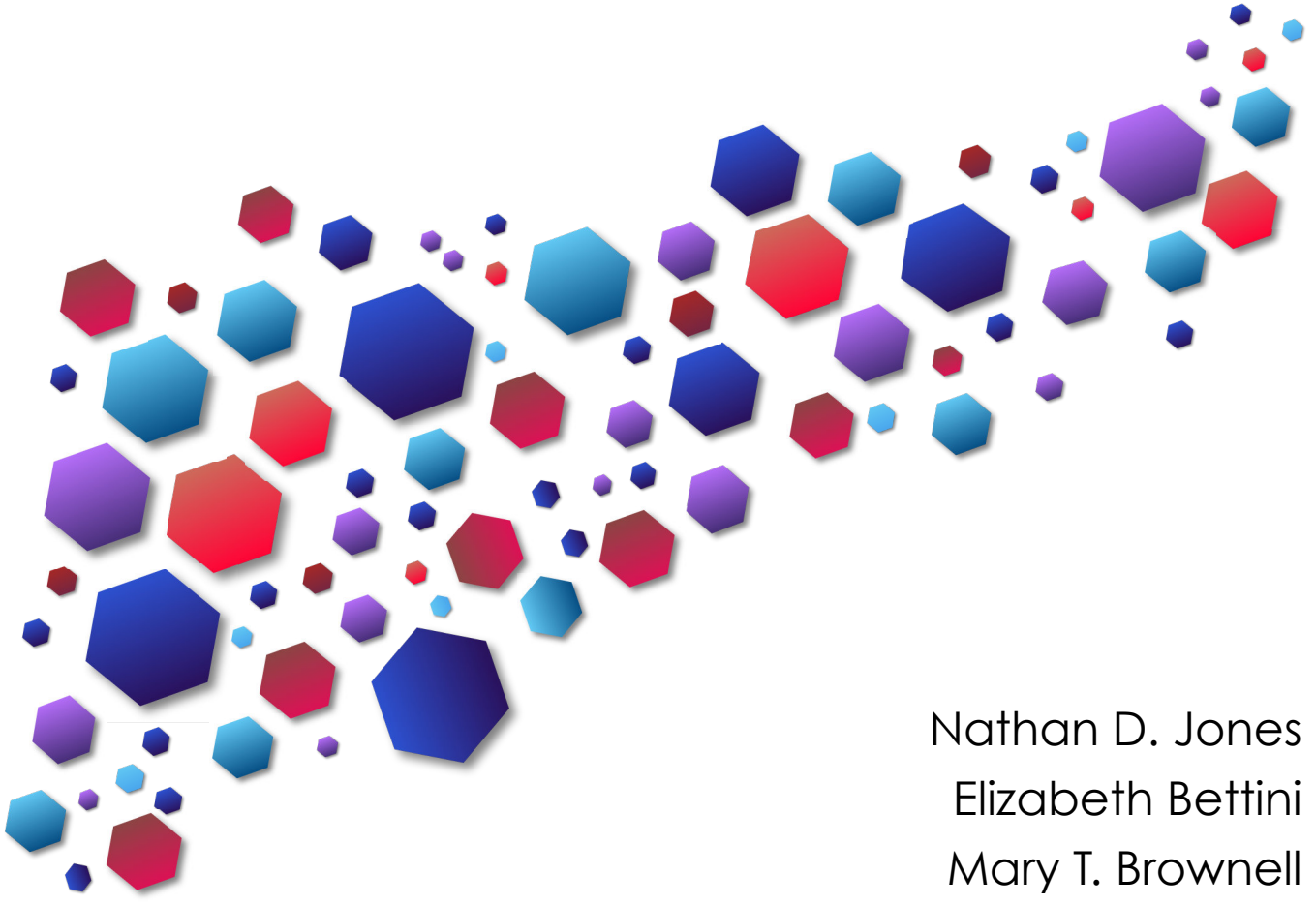
COMPETING STRANDS OF EDUCATIONAL REFORM POLICY

Can Collaborative School Reform and Teacher Evaluation Reform Be Reconciled?



Nathan D. Jones
Elizabeth Bettini
Mary T. Brownell

ALBERT SHANKER INSTITUTE

COMPETING STRANDS OF EDUCATIONAL REFORM POLICY

Can Collaborative School Reform and Teacher Evaluation Reform Be Reconciled?

Nathan D. Jones
BOSTON UNIVERSITY

Elizabeth Bettini
BOSTON UNIVERSITY

Mary T. Brownell
UNIVERSITY OF FLORIDA

April 2016

THE ALBERT SHANKER INSTITUTE, endowed by the American Federation of Teachers and named in honor of its late president, is a nonprofit, nonpartisan organization dedicated to three themes—children's education, unions as advocates for quality, and both civic education and freedom of association in the public life of democracies. Its mission is to generate ideas, foster candid exchanges and promote constructive policy proposals related to these issues.

The institute commissions original analyses, organizes seminars, sponsors publications and subsidizes selected projects. Its independent board of directors is composed of educators, business representatives, labor leaders, academics and public policy analysts.

EXECUTIVE SUMMARY

Both collaborative school reform and teacher evaluation reform show promise for improving teaching and learning. In recent years, U.S. education policy has been focused on improving teacher evaluation, often layering individual teacher evaluation strategies on top of school organizations that value collaborative teaching efforts. These reform efforts have largely been implemented separately from one another, and they are predicated on very different assumptions about how schools and teachers contribute to student learning. Embedded in policies promoting collaborative school structures is the assumption that student learning occurs within a complex social system (the school) that influences teachers and students. As a result, these policies are concerned with changing school organization and systems to foster the kinds of interactions that are likely to improve teaching and learning. In contrast, policies that promote teacher evaluation systems reflect an assumption that individual teachers—whose effectiveness is viewed as stable, portable and independent of context—operating within their own classrooms, are the primary agents capable of and responsible for influencing student learning. Thus, these policies focus on changing schools by improving the average effectiveness of individual teachers, while leaving school systems (e.g., the norms and structures by which educators interact) largely unchanged.

Both policy approaches show promise. Yet, when enacted in schools, they have the potential to undermine one another. Therefore, the purpose of this paper is to begin conceptualizing one avenue for reconciling these policies. We consider how future research might begin to develop measures to evaluate collaborative teaching, specifically teachers' performance as a team engaged in collaboratively enhancing students' learning, through, for instance, co-teaching, tiered instruction, and other models of collaborative teaching.

We review research on the evaluation of teamwork in education and in other fields, such as the military. We find:

- Educational research has yet to systematically investigate how to measure team teaching within schools.
- But teamwork can be measured, and measured rigorously. Education should look to other fields – such as the military – that have successfully developed assessments of team performance.
- Measures of team planning and team knowledge may prove particularly useful and could gradually be integrated within evaluation and accountability systems that support professional learning and ensure that collaborative practices are rewarded.

The recent push for individual teacher accountability raises questions about how teacher evaluation policies can be pursued in ways that support and enhance, rather than undermine, educational policies and practices that promote collaborative teaching practices, such as tiered instruction and co-teaching. We argue that one strategy for addressing the tension between these two strands of research is to conduct further research on assessments of team performance in schools, specifically teachers' collaborative efforts to contribute to student learning. Current evaluation reforms are in their nascence, and their emphasis on identifying effective teachers may have a positive effect on students' learning. But the likelihood that these policies will take hold and resonate with teachers will depend on whether they map onto the realities of teachers' work. Team work is important to most teachers' efforts to meet college and career ready standards, teach higher level skills, and serve the most vulnerable populations of students, such as those with disabilities and English language learners. More effectively evaluating team work in education could be one avenue for reconciling teacher evaluation reform policy with collaborative school reforms. If the two strands of policy reforms are to coexist, we urge policymakers, researchers and practitioners alike to take seriously the need to make space in the evaluation process for the assessment of team work in schools.

COMPETING STRANDS OF EDUCATIONAL REFORM POLICY:

Can Collaborative School Reform and Teacher Evaluation Reform Be Reconciled?

Over the past 20 years, a range of educational policies and practices have promoted more collaborative ways of planning for and providing instruction. These policies and practices include such initiatives as: (a) professional learning communities (PLCs), in which teachers engage in collaborative efforts to learn and improve their instruction (e.g., Giles & Hargreaves, 2006; Newmann & Wehlage, 1995; Vescio, Ross, & Adams, 2008); (b) lesson study, in which groups of teachers engage in professional learning experiences while collaboratively learning about content, planning, teaching, and evaluating their instruction (e.g., Benedict, 2014; Gersten et al., 2009; Lewis, Perry, & Murata, 2006); (c) multi-tiered systems of support (e.g., Response to Intervention, Positive Behavioral Interventions and Supports), in which general educators and specialists collaborate with one another to evaluate all students' needs, provide evidence-based core instruction to all students, and provide more intensive, supplementary instruction to students who struggle (e.g., Coffey & Horner, 2012; Fuchs, Fuchs, & Compton, 2010; Fuchs, Fuchs, & Stecker, 2012; Sugai & Horner, 2006; Sugai, Simonsen, & Horner, 2008); and (d) co-teaching, in which two or more teachers, often a special education teacher and general education teacher, collaboratively plan and provide instruction in the same content and the same setting (e.g., Cook & Friend, 1995; Murawski & Swanson, 2001; Scruggs, Mastropieri, & McDuffie, 2007).

These models all assume that teachers who collaborate together to plan and provide instruction can contribute more to students' learning than one teacher alone. It could be, for example, that in collaborating with colleagues, teachers gain access to the knowledge and resources embedded in their networks of peers; such logic is supported with empirical evidence on how information gained from professional development gets diffused through school organizations (e.g., Frank, Zhao, & Borman, 2004; Sun, Frank, Penuel, & Kim, 2013). Further, there is some evidence that, through positive interactions with colleagues and the potential peer learning that ensues, teachers have greater success in promoting student achievement (e.g., Jackson & Bruegmann, 2009; Kraft & Papay, 2014; Pil & Leana, 2009; Ronfeldt, Framer, McQueen, & Grissom, 2015; Vescio et al., 2008). Kraft and Papay, for example, found that teachers in one North Carolina school district experienced significantly greater gains in their effectiveness over their first 10 years of teaching when they worked in schools with more supportive professional environments, and especially when their schools promoted more peer collaboration. Similarly, Ronfeldt and colleagues' analysis of data from a Florida county found that schools with stronger collaborative cultures elicited significantly stronger student achievement gains than did schools with weaker collaborative cultures, and that teachers working in more collaborative cultures also improved their individual effectiveness more rapidly than teachers working in less collaborative cultures.

While collaborative organizational structures are evident in a number of professional development and school improvement efforts, the primary emphasis of recent state and local education policies has been on evaluating the effectiveness of individual teachers. Government and private sector initiatives have been introduced in the hope of improving teacher evaluation (e.g., Bill & Melinda Gates Foundation, 2010; U.S. Department of Education, 2010), and numerous research efforts have attempted to establish the validity of various measures of teacher effectiveness, including measures of teachers' classroom practices and measures of student growth (e.g., Bell et al., 2012; Casabianca, Lockwood, & McCaffrey, 2014; Ehlert, Koedel, Parsons, & Podgursky, 2012; Hill, Kapitula, & Umland, 2011; Lockwood et al., 2007; Whitehurst, Chingos, & Lindquest, 2014).

Efforts to evaluate individual teachers are intended to strengthen districts' abilities to improve teaching by identifying and promoting effective teachers, while also providing targeted professional development to support teachers in improving their knowledge, skills, and practices (e.g., Kane, Kerr, & Pianta, 2014). Indeed, there is some limited evidence that comprehensive teacher evaluation systems may contribute to improvements in teacher effectiveness over time (Taylor & Tyler, 2012). Analyzing data from Cincinnati public schools, Taylor and Tyler found that teachers were more effective during years in which they were evaluated by well-trained observers, and that their effectiveness improved significantly both initially and in subsequent years.

These two strands of educational policy remain prominent in current reform efforts, but they have largely been divorced from one another, working off of very different assumptions about how schools and teachers contribute to student learning. Embedded in policies promoting collaborative school structures is the assumption that student learning occurs within a complex social system (the school) that influences teachers and students; as a result, these policies are concerned with changing school organization and systems to foster the kinds of interactions likely to improve teaching and learning. In contrast, policies promoting high-stakes teacher evaluation systems reflect an assumption that individual teachers—whose effectiveness is viewed as stable, portable and independent of context—operating within their own classrooms, are the primary agents capable of and responsible for influencing student learning; thus, these policies focus on changing schools by improving the average effectiveness of individual teachers (through promoting high-performing teachers and either removing or developing low-performing ones), while leaving school systems (e.g., the norms and structures by which teachers interact) largely unchanged.

When enacted in schools, the policies based on these different assumptions have the potential to undermine one another (Johnson, 2015; Valli, Croninger, & Walters, 2007). Teachers who are evaluated individually may have an incentive to focus on only those students assigned to their own rosters, investing their time and energy in maximizing their own students' growth in their own content area, rather than maximizing the success of all students through collaboration; thus, individual evaluation systems may undermine collaborative norms in schools (Johnson, 2015). Further, in schools that maintain collaborative norms, the validity of individual evaluation systems may be undermined, as teachers are unlikely to have faith in an individually assigned value-added score when they share instructional responsibilities with many other professionals (Valli et al., 2007).

Given that both of these strands of educational reform policy appear to be valuable tools for improving teaching, we need to consider what kinds of changes would be necessary for teacher evaluation policies and collaborative school reform policies to become mutually supportive. Therefore, the purpose of this paper is to conceptualize how future research might begin to develop measures to evaluate teachers' collaborative teaching and shared contribution to students' learning. We first conducted a systematic review of educational research and scholarship that conceptualized measures of collaborative teaching. We then supplemented the limited educational scholarship by examining research from other professions (e.g., the military) that have developed and researched measures to evaluate the performance of teams working on shared tasks.

1. MODELS AND MEASURES OF TEAMWORK IN EDUCATIONAL RESEARCH

We recognize that collaborative school reform efforts can encompass a broad set of strategies and practices. Thus, for the purposes of this literature review, we define "collaborative teaching" as including any of the three forms of collaborative teaching that Valli and colleagues (2007) identified in their investigation of how schools utilize collaboration for instruction. Analyzing class rosters, teacher logs, interviews, and focus group data from teachers in 18 high-achieving, high-poverty schools, the three collaborative instructional structures that were commonly used were: (a) simultaneous instruction, when multiple adults simultaneously provide instruction to a group of students; (b) supplemental instruction, when students who require additional support in reading or math are pulled out of their core class for some part of the day to receive supplemental instructional support from another teacher; and (c)

sequential instruction, when students change classes throughout the year so that they can be grouped homogeneously and flexibly according to their current skills and learning needs (Valli et al., 2007).

We sought measures of collaborative teaching by conducting a literature search in OneSearch, a library tool that simultaneously searches multiple online databases, including EBSCO, JSTOR, ERIC, and others. The terms *evaluate*, *measure*, *teacher evaluation, teacher effects, assess*, and their other conjugations (e.g., *evaluating*) were paired with the terms *co-teach*, *collaboration*, and *team teach*, and their other conjugations (e.g., *co-teaching, collaborative*). The search was limited to peer-reviewed journals. The date range was initially narrowed from 2001 to 2015, but this yielded only three results, so the date range was expanded to include any articles published after 1990; however, no additional articles were located. Because there is so little literature in this area, any article that examined how to evaluate teachers working collaboratively was included, regardless of whether it was an empirical, conceptual, or practitioner-oriented article. Studies that investigated school norms for collaboration (e.g., Kraft & Papay, 2014; Ronfeldt et al., 2015) were excluded, as these studies do not provide insights into how teams of educators working collaboratively might validly be evaluated. We also limited our search to research conducted in the United States. While examples of systems that promote and reward collaborative practice exist in other countries (for a recent review, see Jensen, Sonnemann, Roberts-Hull, & Hunter, 2016), these systems are integrated within schooling structures that are very different from those of U.S. schools, and their applicability to U.S. schools is therefore unclear.

Only three educational studies met the criteria for this review. The first of these, by Hock and Isenberg (2012), investigated different methods by which student growth models could be constructed to account for teachers' shared contributions to students' learning. The two remaining studies—Wilson (2005) and Murawski and Lochner (2011)—created measures to evaluate special and general educators' collaborative teaching in co-taught classrooms. We summarize the potential contributions of each study below.

Hock and Isenberg (2012) examined various strategies for accounting for shared contributions to student academic growth in teacher value-added models. The three methods reviewed were the partial credit method, the teacher team method, and the full roster method. The partial credit method holds teachers accountable for student learning gains in direct proportion to the amount of time a teacher spent with a given child, without accounting for interactions between teachers or team effects on students' learning gains. The teacher team method holds the team collectively accountable for the learning of students who are shared among team members; this can be combined with the partial credit method in some circumstances, such as when a student changes teachers midyear. Finally, the full roster method assigns shared students to each teacher, such that each student's learning gains count fully, but separately, for each teacher. The full roster method cannot be combined with any other method, but it is more robust to diverse teaming arrangements (e.g., team teaching, supplemental teaching) than the teacher team method. Hock and Isenberg applied all three methods to fourth- through eighth-grade student and teacher data from a large district. They found that the partial credit method produced unstable estimates of teacher effects, but was reliable for estimating school effects on students' learning gains. Meanwhile, the teacher team method and the full roster method provided results that were "theoretically related and empirically similar." The advantage of the teacher team method is that it permitted greater flexibility in the ways that teams were assessed, but the full roster method was simpler (and thus more likely to be perceived as valid by teachers), and it was likely more reliable because it maintained more linkages between students and teachers.

Wilson (2005) developed a tool for observing general and special educators in co-taught classrooms. The observation tool's development team—consisting of Wilson and a team of district and school-level administrators—proceeded in four stages. First, they developed criteria for what they thought a "good" co-taught lesson should look like, including such criteria as a positive relationship between teachers and the acknowledgement and accommodation of diverse learning needs. Next, they questioned whether the evaluation of co-teaching required a different perspective than evaluating classes taught by a single teacher. In order to answer this question, they watched videos of co-taught lessons; their observations of these videos led them to conclude that, in co-taught classrooms, observers should pay particular attention to teachers' roles, instructional activities,

and assessment practices. Third, based on these themes and the criteria for a good co-taught lesson, the team developed an observation tool with questions related to each theme. For example, questions about teachers' roles included, "Does each role enhance the learning process?" and questions about instructional activities included, "What strategies/ modifications are being employed to assist struggling students?" Finally, the team assessed the social validity of the resulting observation tool. Supervisors felt that the tool was helpful, but they stated that some of the questions in it would be more or less relevant for different lessons. Because of this, they also felt that pre-conferences would be important, so that teachers and administrators could collaboratively determine which questions the supervisor should focus on. They also felt that, in order to be informed evaluators, they needed more information about the possible roles, strategies, and assessments that would be appropriate for co-teaching. The team therefore decided to continue meeting to create a companion guide clarifying some questions and providing concrete examples for supervisors (Wilson, 2005). No further research on this tool could be located.

Murawski and Lochner (2011) also developed an observation protocol for evaluating co-teaching, based on their own ideas about what constitutes effective co-teaching. Their observation protocol was structured around three questions that the authors recommend observers should ask. The first question, "What do I ask for?" pointed administrators to documentation of effective co-teaching, such as lesson plans, syllabi, modified student work, and collaborative documentation of students' behaviors. The second question, "What do I look for?" directed administrators to attend to the way teachers interact with one another and with students during instruction, with specific criteria related to differentiation of instruction, parity between educators, and the use of multiple co-teaching techniques. Finally, the third question, "What do I listen for?" focused on shared language, the classroom community, questioning strategies, and the inclusion of all students in instruction. Although Murawski and Lochner provided citations to support their conceptualization of effective co-teaching, those citations are in fact primarily references to other practitioner-oriented publications by the two authors and not to research on what constitutes effective co-teaching. Thus, the components of the tool should be viewed with great caution.

CONCLUSIONS ABOUT MEASURES OF COLLABORATIVE TEACHING

In sum, despite increasing commentary (e.g., Johnson, 2015; Sherer & Barmore, 2015) and empirical evidence (e.g., Kraft & Papay, 2014; Ronfeldt, Framer, McQueen, & Grissom, 2015) surrounding the importance of collaborative, schoolwide reform, the field has offered few tools that could be used to evaluate shared practice. Of these, only the efforts of Hock and Isenberg (2012)—who compared strategies of accounting for shared practice in value-added scores—have resulted in recommended strategies that could be implemented in teacher evaluation policies. Further, the research on measuring collaborative teaching that does exist has focused only on one kind of teacher collaboration—the shared instruction of students with disabilities by general educators and special educators. Thus, we turn now to other disciplines to examine other approaches for evaluating collaborative practice.

2. MODELS AND MEASURES OF TEAM EVALUATION FROM OTHER PROFESSIONS

Many fields rely on intensive collaboration among experts; essential work in medicine, the military, business, and other professions often depends on collaboration among expert professionals, and scholars of these professions have developed a range of tools to measure the performance of teams working collaboratively (Salas, Cooke, & Rosen, 2008). In light of limited educational research examining measures of collaborative teaching, we sought measures of team effectiveness from other fields, to determine how these measures might be adapted to evaluating collaborative teaching.

We applied specific criteria to determine whether measures might be appropriate for evaluating collaborative teaching. First, we sought measures designed to assess teamwork that is similar in complexity to collaborative teaching. Collaboration among teachers is typically either: (a) *reciprocal teamwork*, in which team members coordinate activities that are completed independently, as when reading specialists or special educators provide intensive reading instruction to

struggling readers while general educators provide those same students with core reading instruction; or (b) *intensive teamwork*, in which team members engage in reciprocal, dynamic, and simultaneous interactions among team members, as in the case of co-teaching (for a thorough discussion of team workflow complexity, see Tesluk et al., 1997; for a description of these methods of collaboration among teachers, see Valli et al., 2007). Therefore, measures were examined only if they were designed to assess either reciprocal or intensive workflow. Second, measures were selected for examination only if they also had strong validity evidence, demonstrating consistent relationships with outcomes of team performance. Third, many aspects of teams (e.g., team members' attitudes, beliefs, and knowledge; team members' personal relationships with one another; teams' interaction patterns; the supports provided to teams; the demands made on teams) have been predictive of team success (Mathieu, Maynard, Rapp, & Gibson, 2007; Salas et al., 2008). However, to be consistent with the purposes of teacher evaluation—which serves both improvement and evaluative purposes—measures were included in this review only if they yield information that can be used to provide formative feedback to teams or if they can be used to develop professional learning experiences that support team improvement.

In the following sections, we first provide an overview of how scholars of team performance conceptualize teamwork. We then provide brief descriptions of three intriguing measures that show promise in educational settings.

INTRODUCTION TO THE SCHOLARSHIP OF TEAM PERFORMANCE

Scholars of team performance have conceptualized teamwork as "two or more interdependent individuals that adaptively interact to reach a common goal" (Cooke, Gorman, Myers, & Duran, 2013, p. 256). Teams are treated as information processing units that gather, store, retrieve, and manipulate information within set parameters (i.e., the situation in which the team is operating) to achieve a shared goal (Salas, Cooke, & Rosen, 2008).

In common conceptualizations of teamwork, each individual within a team has four different kinds of knowledge (Cooke et al., 2004; Salas et al., 2008). *Task work* knowledge includes knowledge about one's own task and the team's task. *Teamwork* knowledge is knowledge about team members' roles. Task work and teamwork knowledge include *long-term* knowledge about consistent features of the task and the team. In addition, they also both include *holistic situational* knowledge, or knowledge of the evolving situation within which the team is operating (Cooke et al., 2004; Salas et al., 2008; Cooke et al., 2013). One individual's long-term task work and teamwork knowledge is also often referred to as his or her mental model of the task and the team (Salas et al., 2008). Each of these kinds of knowledge exist at the individual and collective level. And, they exist at a holistic level, which is "team member knowledge that has been processed or integrated … through team behaviors" (Cooke et al., 2004, p. 6; Cooke et al., 2013; Salas et al., 2008).

Holistic knowledge may be more or less than the sum of individual team members' knowledge, depending on the effectiveness of those interactions. For instance, in a highly functioning team that communicates efficiently and effectively, team members may interact in synergistic ways that produce new knowledge or novel solutions, which no single member could have done alone. On the other hand, in a poorly functioning team, team knowledge may be less than the sum of individual members' knowledge, if team processes do not provide opportunities for productive exchanges of relevant information (Cooke et al., 2013). Team cognition is thus an activity, comprised of interactions among team members, and is the focus of the most intriguing and well-validated measures of team performance (Cooke et al., 2013; Salas et al., 2009).

Below, we describe three constructs related to team performance that could be assessed in the work of teachers: (a) holistic situational knowledge; (b) team planning processes; and (c) shared mental models. For each of the three constructs, we first present a definition, then describe how it has been assessed in other fields, and then finally discuss the ways in which it may be applicable to evaluating teachers' teamwork.

A. Assessing Holistic Situational Knowledge

A team's holistic situational awareness—team members' knowledge of the evolving situation in which they are working—is highly predictive of the team's success in achieving its desired outcomes, more predictive than any other measures developed by scholars of team performance (Cooke et al., 2001; Cooke et al., 2004; Lee, Kim, & Seong, 2012). To assess holistic situational awareness, Cooke and colleagues (2001) queried each team member at three random, unspecified times during the completion of a simulated aviation task. Queries asked about the evolving situation of the mission, including questions related to what had been accomplished, what they should do next, and what they predicted would be accomplished by the end. Team members' collective accuracy in answering these questions significantly predicted teams' effectiveness at accomplishing their mission (Cooke et al., 2001).

Application to evaluating collaborative teaching.

Holistic situational awareness is analogous to the pedagogical concept of "with-it-ness" (Brophy & Goode, 1986), the idea that teachers must be aware of and sensitive to the complex, shifting interactions constantly occurring between and among students and instructional activities within their classrooms. This type of assessment could be used to evaluate co-teachers' holistic situational awareness (or "with-it-ness") during co-taught instruction. For instance, co-teachers could be queried about students' learning during collaborative instructional activities. The specific questions would likely need to vary depending on the nature of the instruction, but could include questions such as:

- What proportion of students have mastered this concept?
- Which students still need support to master this concept?
- Which students do you anticipate will not master this concept today?
- Which students are currently off task?

To validate the measure, co-teachers' collective accuracy in answering these questions could then be correlated with objective measures of students' learning.

Such a measure does present several challenges, however. First, to determine the accuracy of teachers' answers, an objective answer to those questions would need to be obtained. To answer the first three questions, an objective measure of student mastery could be integrated into the lesson at the same time point as the query. Answering the fourth question would likely rely on observer judgment. Second, the specific questions asked would likely need to vary depending on the lesson. This method would therefore require evaluators to engage in sophisticated pre-observation interactions with teachers, to develop the necessary questions and measures. Because of the sophisticated knowledge and planning involved, this measure may be most useful for research, rather than practice, at least for the present.

B. Assessing Team Planning Processes

The quality of teams' planning significantly predicts outcomes of teamwork (Mathieu & Schulze, 2006; Mathieu & Rapp, 2009; Weil, Freeman, Carley, Cooke, & Diesner, 2006). Mathieu and Rapp found that, on a business simulation task, teams that began their team activities by developing plans that thoroughly and consistently articulated team members' roles, their processes for completing work together, and their plans for managing conflict outperformed teams whose charters were less thorough or internally inconsistent. Teams' articulation of their strategies during pre-planning also significantly predicted their success on this task; teams whose strategic planning thoroughly and coherently addressed all aspects of their plan, including contingencies for unexpected events, outperformed teams whose strategic planning was less thorough and less internally consistent (Mathieu & Rapp, 2009). Mathieu and Schulze and Mathieu and Rapp assessed teams' planning using expert ratings of formal documentation of teams' plans (i.e., team charters, strategic planning documents).

Application to evaluating collaborative teaching.

Evaluations of team teachers' lesson planning documentation could prove useful. Researchers could

examine the elements of team lesson plans that are associated with positive student learning outcomes and then develop expert rating systems that school administrators, coaches, and professional development providers could use to provide teams with formative feedback about how to develop plans that more thoroughly and consistently address all students' learning needs, and to clearly articulate meaningful roles for all team members.

As an example, because a successful Response to Intervention framework depends on collaboration across general education and special education teachers, lesson plans could be evaluated for the degree to which they plan for purposeful, high-quality, well-aligned instruction. The criteria used to assess collaborative planning would likely need to vary to reflect the nature of instruction at each tier of RTI; for instance, at Tier 1, lesson planning documents would be expected to include the accommodations and modifications being provided to ensure all students have access to the content, whereas lesson planning documents at higher tiers would need to identify the ways in which content is aligned with the Tier 1 instructional goals while remediating skill deficits for struggling learners.

C. Assessing Shared Mental Models

The term "mental models" refers to schemas, or underlying structures of each team member's knowledge about the task, the team, the situation, and the long-term goals (Kraiger & Wenzel, 1997; Matheiu et al., 2000). Mental models explain relationships among constructs and stimuli, and thus allow individuals to interpret their environment, make predictions, and act effectively (Kraiger & Wenzel, 1997; Matheiu et al., 2000). When team members share common mental models, they have a common understanding of the environment, and thus may be better able to accurately engage in coordinated interactions within a dynamic situation (Banks & Millward, 2007; Mathieu et al., 2000). A number of studies have found that the degree to which team members' mental models are shared and accurate predicts positive outcomes of teamwork (Banks & Millward, 2007; Edwards, Day, Arthur, & Bell, 2006; Mathieu et al., 2000). The development of shared mental models relies upon feedback; teams that do not receive feedback on their performance do not seem to develop shared mental models (Matheiu et al., 2000).

It is important to note that shared mental models are not required for every aspect of a task (Cooke et al., 2013; Weil, Freeman, Carley, Cooke, & Diesner, 2006). When team members have highly differentiated roles, some of their knowledge will necessarily be dissimilar. In these cases, they may need to share some core knowledge (e.g., the goal, the general parameters of the task, the materials available for completing the task, the roles of the different team members; Weil et al., 2006), but for other knowledge (e.g., knowledge of their specialized roles), shared mental models are less important, provided that team interaction processes promote efficient and constructive exchanges among team members in specialized roles (Cooke et al., 2013). Mental models may need to be most similar when team members' roles within a task are relatively similar (e.g., a team of third-grade teachers working together to develop a lesson during a lesson-study cycle; Cooke et al., 2013).

Several methods have been developed to assess shared mental models (Mohammed et al., 2000; Weil et al., 2006). In two methods (path finder analysis and multidimensional scaling), researchers conduct a task analysis to select key concepts, which team members then rate for their relatedness (e.g., Banks & Millward, 2007; Mathieu et al., 2000; Mohammed et al., 2000). In other methods (text-based causal mapping, interactively elicited causal mapping, network text analysis), team members' natural or elicited talk is analyzed for key concepts. In all methods, algorithms then evaluate the relatedness of concepts and create a network structure in which nodes represent key concepts, and the lines connecting nodes represent the strength of the relationships among concepts (Mohammed et al., 2000). Team members' matrices are assessed for their similarity to one another and their similarity to an expert's network, which is used as a measure of accuracy (Banks & Millward, 2007; Mohammed et al., 2000). Individual accuracy, team accuracy, and shared team networks all predict a team's success at achieving its goals (Banks & Millward, 2007; Mohammed et al., 2000).

Application to evaluating collaborative teaching.

This form of assessment could be useful for informing the development of professional learning experiences. Since feedback seems to be especially important for the development of shared mental mod-

els (Matheiu et al., 2000), providers of professional development could use this type of assessment to determine what aspects of teaching teams' mental models are divergent or inaccurate. They could then focus their feedback to the team on these specific areas, to support team members in developing a shared and accurate understanding of those particular aspects of their classroom and their instruction.

CONCLUSIONS

The recent push for individual teacher accountability raises questions about how these policies can be pursued in ways that support and enhance, rather than undermine, educational policies and practices that promote effective collaboration. We argue that one strategy for addressing the tension between these two strands of research is to conduct further research on assessments of collaboration. Other fields—such as surgery and the military—have developed rigorous measures of teamwork, which provide a starting point for similar assessment development in education. When examining the three constructs related to team performance (holistic situational knowledge, team planning processes, and shared mental models), cases can be made about the viability of each as applied to teaching.

With respect to the assessment of team planning processes, there is already precedent for evaluation practices that ask administrators to rate a teacher's effectiveness in his or her planning and preparation; these would need to be modified to account for alignment across individual teachers' lesson plans. An assessment of teams' holistic situational knowledge seems to be promising, and it is intuitively a critical aspect of teachers' shared practice; for example, it is easy to imagine that a team's effectiveness depends on team members' ability to "read the room" and adapt instruction on the fly. At the same time, such an assessment would require more investment of a school's time and resources, as "true scores" of such things as student engagement or academic performance would be necessary to collect in the moment. Finally, assessing shared mental models (i.e., team members' common understanding of their task and the work required of each team member) would not be resource intensive and would help to focus professional development efforts.

In sum, while further measurement research will be necessary to incorporate measures of any of the above constructs into teacher evaluation, other fields have provided an empirical base for how the field of education might rigorously assess teamwork. It is not difficult to imagine including the assessment of team planning into observational measures or to include the assessment of shared mental models into teacher instructional surveys. Measuring holistic situational knowledge would require more development work than the other two constructs, but it also has an intuitive appeal that seems to get at one of the core aspects of effective teaching. Current evaluation reforms are in their nascence, and their emphasis on identifying effective teachers will hopefully have a positive effect on students' learning. But the likelihood that these policies will take hold and resonate with teachers will depend on whether they map onto the realities of teachers' work. Collaborative practice is important to most teachers' practice and is essential for the instruction of many student populations, including those with disabilities. Thus, if the two strands of policy reforms are to coexist, we urge researchers and practitioners alike to take seriously the need to make space in the evaluation process for the assessment of teamwork.

REFERENCES

Banks, A. P. & Millward, L. J. (2007). Differentiating knowledge in teams: The effect of shared declarative and procedural knowledge on team performance. *Group Dynamics: Theory, Research, and Practice, 11,* 95-206.

Bell, C., Gitomer, D., McCaffrey, D., Hamre, B., Pianta, R., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment, 17.*

Benedict, A. E. (2014). *Learning together: General and special education teachers' evolving understandings of word study content during ongoing collaborative professional development* (unpublished doctoral dissertation). Gainesville: University of Florida.

Bill & Melinda Gates Foundation (2010). Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project. MET Project Research Paper. Seattle, WA: Bill & Melinda Gates Foundation.

Brophy, J. & Goode, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), Third handbook of research on teaching (pp. 328-375). New York: American Educational Research Association.

Casabianca, J. M., Lockwood, J. R., & McCaffrey, D. F. (2014). Trends in classroom observation scores. *Educational and Psychological Measurement.* Advance online publication. doi: 10.1177/0013164414539163.

Coffey, J. H. & Horner, R. H. (2012). The sustainability of schoolwide positive behavior interventions and supports. *Exceptional Children, 78,* 407-422.

Cook, L. & Friend, M. (2010). The state of the art of collaboration on behalf of students with disabilities. *Journal of Educational and Psychological Consultation, 20,* 1-8.

Cooke, N. J., Kiekel, P. A., & Helm, E. E. (2001). Measuring team knowledge during skill acquisition of a complex task. *International Journal of Cognitive Ergonomics, 5,* 297-315.

Cooke, N. J., Salas, E., Kiekel, P. A., & Bell, B. (2004). Advances in measuring team cognition. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 83-106). Washington, DC: American Psychological Association.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science, 37,* 255-285.

Edwards, B. D., Day, E. A., Arthur, W., & Bell, S. T. (2006). Relationships among team ability composition, team mental models, and team performance. *Journal of Applied Psychology, 91,* 727-736.

Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. (2012). Selecting growth measures for school and teacher evaluations. National Center for Analysis of Longitudinal Data in Education Research (CALDAR). Working Paper #80.

Frank, K. A., Zhao, Y., & Borman, K. (2004). Social capital and the diffusion of innovations within organizations: The case of computer technology in schools. *Sociology of Education*, *77*(2), 148-171.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2010). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78,* 263-279.

Fuchs, D., Fuchs, L. S., & Stecker, P. M. (2012). The "blurring" of special education in a new continuum of general education placements and services. *Exceptional Children, 76,* 301-323.

Gersten, R., Dimino, J., Jayanthi, M., Kim, J. S., & Santoro, L. E. (2010). Teacher study group: Impact of the professional development model on reading instruction and student outcomes in first grade classrooms. *American Educational Research Journal, 47,* 694-739.

Giles, C. & Hargreaves, A. (2006). The sustainability of innovative schools as learning organizations and professional learning communities during standardized reform. *Educational Administration Quarterly, 42,* 124-156.

Hill, H., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal, 48*, 794-831.

Hock, H. & Isenberg, E. (2012). Methods for accounting for co-teaching in value added models. *IDEAS Working Paper Series from RePEc.* Retrieved from: http://ideas.repec.org/p/mpr/mprres/e53ebf9f792e48dca345a4bba26a1d70.html.

Jackson, C. K. & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics, 1,* 85-108.

Jensen, B., Sonnemann, J., Roberts-Hull, K., Hunter, A. (2016). Beyond PD: Teacher professional learning in high-performing systems. Washington DC: National Center on Education and the Economy.

Johnson, S. M. (2015). Will VAMs reinforce the walls of the egg-crate school? *Educational Researcher, 44,* 117-126.

Johnson, S. M., Papay, J. P., Fiarman, S. E., Munger, M. S., & Qazilbash, E. K. (2010). Teacher to teacher: Realizing the potential of peer assistance and review. *Washington, DC: Center for American Progress*.

Kane, T., Kerr, K., & Pianta, R. (2014). *Designing teacher evaluation systems: New guidance from the measures of effective teaching project*. John Wiley & Sons

Kraft, M. A. & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis, 36,* 476-500.

Kraiger, K., & Wenzel, L. 1997. Conceptual development and empirical evaluation of measures of shared mental models as indicators of team effectiveness. In M. Brannick, E. Salas, & C. Prince (Eds.), *Team Performance Assessment and Measurement*: 63-84. Hillsdale, NJ: Lawrence Erlbaum.

Lee, S. W., Kim, A. R., & Seong, P. H. (2012). Measuring situation awareness of operation teams in NPPs using a verbal protocol analysis. *Annals of Nuclear Energy, 43,* 167-175.

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational researcher*, *35*(3), 3-14.

Lockwood, J., McCaffrey, D., Hamilton, L., Stecher, B., Le, V., & Martinez, J. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement, 44*, 47-67.

Matheiu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85,* 273-283.

Mathieu, J. E. & Rapp, T. L. (2009). Laying the foundation for successful team performance trajectories: The roles of team charters and performance strategies. *Journal of Applied Psychology, 94,* 90-103.

Mathieu, J. E. & Schulze, W. (2006). The influence of team knowledge and formal plans on episodic team process-performance relationships. *The Academy of Management Journal, 49,* 605 – 619.

Mathieu, J., Maynard, M. T., Rapp, T., & Gibson, L. (2007). Team effectiveness 1997 – 2007: A review of recent advancements and a glimpse into the future. *Journal of Management, 34,* 410-476.

Mohammed, S., Klimoski, R., & Rentsch, J. R. (2000). The measurement of team mental models: We have no shared schema. *Organizational Research Methods, 3,* 123-165.

Murawski, W. W. & Swanson, L. (2001). A meta-analysis of co-teaching research: Where are the data? *Remedial and Special Education, 22,* 258-267.

Murawski, W. W. & Lochner, W. W. (2011). Observing co-teaching: What to ask for, look for, and listen for. *Intervention in School and Clinic, 46,* 174-183.

Newmann, F., & Wehlage, G. (1995). Successful School Restructuring. Madison, WI: Center on Organization and Restructuring of Schools.

Pil, F. K., & Leana, C. (2009). Applying organizational research to public school reform: The effects of teacher human and social capital on student performance. *Academy of Management Journal*, *52*, 1101-1124.

Ronfeldt, M., Framer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal, 52,* 475-514.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors, 50,* 540-547.

Scruggs, T., Mastropieri, M. A., & McDuffie, K. A. (2007). Co-teaching in inclusive classrooms: A metasynthesis of qualitative research. *Exceptional Children, 73,* 392-416.

Sherer, D., & Barmore, J. (2015). What makes teacher collaboration work? *Shanker Blog.* Retrieved from: www.shankerinstitute.org/blog/what-makes-teacher-collaboration-work.

Sugai, G. & Horner, R. R. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School Psychology Review, 35,* 245-259.

Sugai, G., Simonsen, B., & Horner, R. H. (2008). School-wide positive behavior supports: A continuum of positive behavior supports for all students. *Teaching Exceptional Children, 40,* 5.

Sun M., Penuel W. R., Frank K. A., Gallagher H. A., & Youngs P. (2013). Shaping professional development to promote the diffusion of instructional expertise among teachers. *Educational Evaluation and Policy Analysis, 35*, 344-369.

Taylor, E. S. & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *The American Economic Review, 102,* 3628-3651.

Tesluk, P., Matheiu, J. E., Zaccaro, S. J., & Marks, M. (1997). Task and aggregation issues in the analysis and assessment of team performance. In M. T. Brannick, C. Prince, & E. Salas (Eds.), *Team performance assessment and measurement* (pp. 197-224). Mahweh, NJ: Lawrence Erlbaum.

U. S. Department of Education. (2010). *American Recovery and Reinvestment Act, Section 1512.*

Valli, L., Croninger, R. G., & Walters, K. (2007). Who (else) is the teacher?: Cautionary notes on teacher accountability systems. *American Journal of Education, 113,* 635-662.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24,* 80-91.

Weil, S. A., Freeman, J., Carley, K. M., Cooke, N. J., & Diesner, J. (2006). Measuring situational awareness through analysis of communications: A preliminary exercise. In: Proceedings of the 2006 command and control research and technology symposium, San Diego, CA.

Whitehurst, G.J., Chingos, M.M., & Lindquest, K.M. (2014). *Evaluating teachers with classroom observers: Lessons learned in four district.* Washington, D.C.: Brown Center for Education Policy at Brookings.

Wilson, G. L. (2005). This doesn't look familiar!: A supervisor's guide for observing co-teachers. *Intervention in School and Clinic, 40,* 271-275.

ABOUT THE AUTHORS

NATHAN JONES is an assistant professor of special education at Boston University. In his research, Dr. Jones focuses on teacher quality, teacher development, and school improvement, with a specific emphasis on the use of measures of teacher effectiveness in evaluation systems. He previously worked as a middle school special education teacher in the Mississippi Delta. Dr. Jones is currently the principal investigator of a $1.6 million IES grant examining whether the Framework for Teaching (FFT) – an observation system used in over 20 states and hundreds of school districts – can be used validly and reliably in the evaluation of special education teachers. He is also co-PI on a W.T. Grant study examining Los Angeles Unified School District's (LAUSD) three-year roll-out of a consequential teacher evaluation system. Dr. Jones currently serves an an editor of The Elementary School Journal and is on the editorial board at The Journal of Teacher Education.

ELIZABETH BETTINI is an Assistant Professor in the Special Education program in the School of Education. At Boston University. She previously taught students with EBD in self-contained and inclusive settings in Arizona and Connecticut. Her research examines how working conditions influence special educators' efforts to effectively serve students with disabilities, especially students with emotional and behavioral disabilities (EBD). Currently, she is working on several research projects investigating the consequences and predictors of novice special educators' perceptions that their workloads are manageable.

MARY BROWNELL is a Professor of Special Education at the University of Florida and Director of the Collaboration for Effective Educator Development, Accountability and Reform (CEEDAR Center), an OSEP funded project designed to improve the preparation of teachers and leaders working with students with disabilities. Dr. Brownell has secured multiple grants (totaling over 42 million dollars) from the U.S. Department of Education (U.S. DOE), Office of Special Education Programs and Institute for Education Sciences to support research and technical assistance aimed at improving teaching quality for students with disabilities. Specifically, she has studied issues related to teacher education, teacher assessment, professional development, and teacher attrition. Dr. Brownell's research efforts have culminated in numerous refereed publications, book chapters, books, and presentations. She and her colleagues edited the first Handbook of Research on Special Education Teacher Preparation, and she recently authored a chapter for the prestigious Handbook for Research on Teaching, published by the American Educational Research Association. She has been recognized nationally for her leadership and research, winning the TED Pearson Award for Excellence in Teacher Education, the Division for Research Hallahan and Kauffman Distinguished Researcher Award, and the David Imig Teacher Achievement Award for lifetime achievement in teacher education.

ALBERT SHANKER INSTITUTE

555 New Jersey Ave., NW | Washington, D.C. 20001
Tel: 202.879.4401 | shankerinstitute.org