DESIGNING TEACHER EVALUATION SYSTEMS

NEW GUIDANCE *from the* MEASURES OF EFFECTIVE TEACHING PROJECT

THOMAS J. KANE
KERRI A. KERR
ROBERT C. PIANTA
EDITORS

BILL & MELINDA GATES *foundation*

JOSSEY-BASS™
A Wiley Brand

DESIGNING TEACHER EVALUATION SYSTEMS

New Guidance from the Measures of Effective Teaching Project

THOMAS J. KANE
KERRI A. KERR
ROBERT C. PIANTA
EDITORS

JB JOSSEY-BASS™
A Wiley Brand

ABOUT THE BILL & MELINDA GATES FOUNDATION

Guided by the belief that every life has equal value, the Bill & Melinda Gates Foundation works to help all people lead healthy, productive lives. In developing countries, it focuses on improving people's health with vaccines and other life-saving tools and giving them the chance to lift themselves out of hunger and extreme poverty. In the United States, it seeks to significantly improve education so that all young people have the opportunity to reach their full potential. Based in Seattle, Washington, the foundation is led by CEO Jeff Raikes and co-chair William H. Gates, Sr., under the direction of Bill and Melinda Gates and Warren Buffett.

BILL & MELINDA GATES foundation

CHAPTER 4

How Framework for Teaching and Tripod 7Cs Evidence Distinguish Key Components of Effective Teaching

RONALD F. FERGUSON WITH CHARLOTTE DANIELSON

ABSTRACT

This chapter uses data from the Measures of Effective Teaching project to study ways that adult observations using the Framework for Teaching and student perceptions using Tripod survey assessments help distinguish components of effective teaching. The approaches are found to be compatible in the components of teaching that they measure. Moreover, adults and students evaluate teaching similarly. The chapter uses value-added test score measures and student survey responses for happiness in class, effort in class, and whether the teacher inspires an interest in college as key outcomes predicted by teaching quality. The mix of teaching components that predicts value added differs systematically from the combination that predicts happiness, effort, or inspiration.

INTRODUCTION

This chapter distinguishes multiple aspects of teaching that together predict student engagement and learning. Our primary aim is to help elementary and secondary school educators understand the components of teaching effectiveness—the types of action that produce or facilitate learning and healthy development—in order to more strategically and effectively improve their own and others' teaching.

It is well known from research that some teachers routinely produce more learning than others (e.g., Kane, McCaffrey, & Staiger, 2010, 2012; Rivkin, Hanushek, & Kain, 2005). Why? According to focus-group research with regular citizens, people believe the reason is that effective teachers simply care more (Chart with Kendall-Taylor, 2008). Similarly, when we ask large audiences of professional educators to select among multiple reasons that some teachers produce more learning, they too select *caring* as the most important reason. Are they correct? Based on classrooms sampled from more than two hundred schools in six cities that participated in the Bill & Melinda Gates Foundation project on Measures of Effective Teaching (MET), findings in this chapter indicate that caring is the strongest predictor of happiness, but not learning. Instead, we find that *classroom management* is the strongest predictor of learning. The chapter presents this and other findings, distinguishing among multiple aspects of teaching and their implications for students.

Three rapidly spreading methods for assessing teacher performance are classroom observations, student surveys, and test-based measures of student learning (i.e., growth or value-added scores). The latter—test-based measures—can help us understand how much students have learned. However, they do not indicate which aspects of teaching may need to improve in order that students might learn more in any particular classroom.

Accordingly, this chapter applies two popular frameworks and associated assessment tools for measuring what teachers actually do in their classrooms. A central question is whether these two approaches—Charlotte Danielson's Framework for Teaching (FfT) and Ronald Ferguson's 7Cs framework from his Tripod Project survey assessments—are mutually reinforcing as ways of diagnosing teachers' professional strengths along with areas in need of improvement. The approaches were developed independently by the authors of this chapter and are widely used in the United States and increasingly abroad. They are research-based and have been refined over more than a decade based on analyses of prior results and feedback from elementary and secondary school practitioners and fellow researchers. Both played central roles in the Bill & Melinda Gates Foundation MET project.

The Framework for Teaching (FfT) is a definition of teaching quality and a classroom observation system designed to enrich deliberations in school systems on ways of improving instruction. Similarly, Tripod survey assessments were designed to measure perceptions of teaching quality and engagement in learning. From the Tripod surveys, MET used the 7Cs framework for effective teaching in addition to a selection of Tripod student engagement survey items that measure happiness in class, effort in class, and whether the teacher inspires students to attend college. We use these measures of happiness, effort, and inspiration to supplement value-added achievement gains as teaching outcome measures. The chapter uses data from fourth through eighth grade classrooms in the six MET districts.

The chapter is focused on three tasks. First, after introducing the frameworks, we demonstrate that specific domains and components of the FfT and Tripod 7Cs frameworks are compatible, not only conceptually, but also empirically. In particular, both frameworks have components that focus on classroom management and others that focus on instruction. Data from these distinct sources provide multiple measures for making judgments about specific categories of teaching practice and can be used together in a coherent program of professional measurement, learning, and support.

Second, we show that having multiple desired outcomes warrants a balance in teaching priorities. This is because the components of the FfT and Tripod 7Cs frameworks that most strongly predict happiness in class and being inspired by the teacher to attend college are different from those that most strongly predict value-added learning gains on standardized exams. On the one hand, the components that most strongly predict value-added achievement gains are associated with keeping students busy and on task and pressing them to think rigorously and persist in the face of difficulty. We refer to these as *press*. On the other hand, components that most strongly predict happiness in class and inspiration to attend college are associated with caring teacher-student relationships, captivating lessons, and other practices that students experience as supportive. We refer to the latter as *support*. Press and support are both important if we care not only about annual test-score gains, but also about the quality of life at school, inspiration to attend college, and a love of learning.

Third, we consider ways that combining and comparing data generated using the two frameworks can contribute to quality control by helping to uncover dishonesty or other problems with implementation. We show that either too much or too little similarity in FfT as compared to Tripod 7Cs data patterns can signal irregularities of implementation or interpretation and may warrant official scrutiny, especially under high-stakes conditions.

Finally, at the end of the chapter, we distill some key implications. Generally, we propose that paying attention to the components of the FfT and Tripod 7Cs frameworks—not just the composite scores—can enrich the quality of reflection, discourse, and support that teachers experience in collaboration with supervisors and peers concerning their teaching. This, in turn, can enhance the quality of instruction that students experience, how hard they work, how much they learn, how happy they are in class, and how earnestly they aspire to attend college.

INTRODUCTION TO THE FRAMEWORKS

Both the FfT and Tripod 7Cs frameworks are multifaceted research-based conceptions of teaching, describing what teachers do in the practice of their profession. The FfT is the outgrowth of Charlotte Danielson's experience at the Educational Testing Service (ETS), where she was a member of the design team for Praxis III (the observation-based system of teacher assessment used for the licensing of beginning teachers) and participated as well in redesigning assessments for the National Board for Professional Teaching Standards. While the FfT emerged from Danielson's work at the national level, the Tripod surveys grew out of Ron Ferguson's work in Shaker Heights, Ohio, and then with a large number of school districts across the country, where the focus was on finding levers to raise achievement levels and narrow achievement gaps. Some states and districts are using observational data from FfT and student survey data from Tripod for both teacher evaluation formulas and professional development planning.

According to MET publications, FfT and Tripod 7Cs measurement tools produce valid and reliable indicators of teaching quality—often more reliable than value added—when administered at the classroom level with fidelity (Cantrell & Kane, 2013; Ho & Kane, 2012; Kane, McCaffrey, & Staiger, 2010, 2012). For readers not familiar with the concept, value added refers to a particular approach to measuring test score gains. What distinguishes value-added measures from simpler test score growth measures is that they are adjusted for between-classroom differences in student characteristics. Many analysts prefer value added for measuring teacher effectiveness because, if implemented properly, value added approximates a condition in which there is no difference across classrooms in the characteristics of the students. Hence, value added for any particular teacher is an estimate of how much that teacher adds to students' skills and knowledge.

When we consider different classrooms taught by the same teachers in the data for this chapter, the between-classroom correlations are 0.38 for value added, 0.42 for the FfT composite, and 0.61 for the Tripod 7Cs composite. Hence, all three metrics are ways of detecting consistency at the teacher level from one classroom to another. Furthermore, as MET reports show, there is cross-validation. Specifically, the fact that each metric is clearly correlated with the others helps validate that all three are indicators of instructional quality.

FRAMEWORK FOR TEACHING

To ensure fidelity of FfT data gathering, the MET project trained hundreds of raters to score video recordings from participating classrooms. (This was done not only for FfT, but also for the other observational protocols that MET employed.) In addition, beyond the MET project, Danielson and colleagues have devised ways of training school-based raters—typically administrators—and then measuring and certifying their rating proficiency. This is important, since the findings in both this chapter and MET are only indicative of what practitioners might find if they use the FfT properly.

Danielson's work on the Praxis III and the National Board assessments at ETS proved important not only because they provided methods of assessing instruction, but more important, because they helped produce standards of practice and concrete guidance for teachers on how to achieve high standards in their classrooms. They provided foundations for teachers to engage in activities that supported teacher learning—self-assessment of teaching skills; reflection on their practice; and professional conversations with peers, coaches, and supervisors. Even in the context of high-stakes assessments of practice, educators found the exercises valuable. The many encouraging responses from educators inspired Danielson to develop the FfT (Danielson, 2013).

In the FfT, the complex activity of teaching is divided into twenty-two components, clustered in four domains of teaching responsibility:

Domain 1: Planning and Preparation

Domain 2: The Classroom Environment

Domain 3: Instruction

Domain 4: Professional Responsibilities

Domains 2 and 3 are the ones that most directly concern the actual delivery of classroom instruction. Each comprises several components. This chapter is focused on four components from Domain 2 (the first four directly below) and four from Domain 3.[1] They are the following:

- *Creating an environment of respect and rapport*, for example: respectful talk, active listening, and turn taking; acknowledgment of students' backgrounds and lives outside the classroom; body language indicative of warmth and caring; physical proximity; politeness and encouragement; and fairness.
- *Establishing a culture for learning*, for example: belief in the value of what is being learned; high expectations, supported through both verbal and nonverbal behaviors, for both learning and participation; expectation of high-quality work on the part of students; expectation and recognition of effort and persistence on the part of students; and high expectations for expression and work products.
- *Managing classroom procedures*, for example: smooth functioning of all routines; little or no loss of instructional time; students playing an important role in carrying out the routines; students knowing what to do and where to move.
- *Managing student behavior*, for example: clear standards of conduct, possibly posted and possibly referred to during a lesson; absence of acrimony between teacher and students concerning behavior; teacher awareness of student conduct, including preventative awareness; absence of misbehavior; and reinforcement of positive behavior.
- *Communicating with students*, for example: clarity of lesson purpose; clear directions and procedures specific to the lesson activities; absence of content errors and clear explanations of concepts and strategies; and correct and imaginative use of language.
- *Using questioning and discussion techniques*, for example: questions of high cognitive challenge, formulated by both students and teacher; questions with multiple correct answers or multiple approaches, even when there is a single correct response; effective use of student responses and ideas; discussion, with the teacher stepping out of the central, mediating role; focus on the reasoning exhibited by students in discussion, both in

give-and-take with the teacher and with their classmates; high levels of student participation in discussion.

- *Engaging students in learning*, for example: students show enthusiasm, interest, thinking, problem solving, etc.; learning tasks that require high-level student thinking and invite students to explain their thinking; students highly motivated to work on all tasks and persist, even when the tasks are challenging; students actively "working," rather than watching while the teacher "works"; and suitable pacing of the lesson, neither dragged out nor rushed, with time for closure and student reflection.
- *Using assessment in instruction*, for example: the teacher paying close attention to evidence of student understanding; the teacher posing specifically created questions to elicit evidence of student understanding; the teacher circulating to monitor student learning and to offer feedback; and students assessing their own work against established criteria.

For the MET project, hundreds of experienced educators were trained to rate classrooms on each of the eight components listed above. They watched video recordings from the MET classrooms and assigned a score of 1, 2, 3, or 4 to each FfT component, representing "unsatisfactory," "basic," "proficient," and "distinguished," respectively. The data for this paper come from the 2009–2010 school year and include an average of two observations per classroom (hence the score for each of the eight components is most often the average from two observations of a classroom). For MET, there were no scores assigned below the component level. However, the score on each component was intended to reflect the rater's judgments concerning the overall performance of the teacher on the elements within that component.

The analysis in this chapter concerns the eight FfT components listed above and the ways that they relate both conceptually and statistically to the 7Cs of the Tripod framework described below.

THE TRIPOD 7Cs MODEL

The Tripod Project emerged in 2000 from a week-long summer workshop that Ron Ferguson designed with educators in Shaker Heights, Ohio. The week focused on Erik Erikson's first five stages of life-cycle identity development. The five clusters of issues, adapted to classrooms, concerned (1) building trusting relationships; (2) cultivating good behavior and cooperation; (3) helping students to set ambitious goals for learning; (4) encouraging and enabling persistence and resilience in the face of difficulty; and (5) helping students develop a sense of academic efficacy and take satisfaction in achievement. Teachers in small groups wrote reports on ways to achieve desired outcomes for each cluster of issues and to avoid their negative opposites. Activities were developed to continue the work during the school year. This included the idea to survey students about their experiences *in particular classrooms* (as opposed to whole-school climate surveys) in order to understand and track progress on instructional improvement.

The concept of the "Tripod Project" developed as a way to cultivate the type of teaching necessary to succeed with the five clusters of issues—later called the "Tripod Engagement Framework"—adapted from the Erikson framework. The "tripod" was "content, pedagogy, and relationships." The idea was that, in order to deliver instruction effectively, teachers needed an understanding of the subjects they were teaching (content knowledge), they needed sufficient skill to help students achieve understanding (pedagogic knowledge and skill), and they needed to connect with students on a personal level so that students would be inspired to trust and cooperate (relationships).

The first few years of Tripod surveys were designed by Ferguson in consultation with Shaker Heights teachers and administrators. They were informed by the interests that the educators expressed, as well as by the research literature on student engagement and teaching practices. Lead teachers helped by reacting to survey drafts and testing their students' interpretations of the items. Initially, there was a survey for grades K through 5 and another for grades 6 through 12. Over the ensuing years, surveys for teachers were developed, and a separate survey was designed for grades K to 2. A clear distinction developed between measures of student engagement (what individual students do, think, and feel) versus student perceptions of teaching (what teachers do and how the classroom operates as measured by the Tripod 7Cs framework).

In 2009, the Bill & Melinda Gates Foundation selected Tripod to supply the student perception surveys for the MET project. By December 2010, MET had produced evidence that the Tripod 7Cs measures are valid and reliable predictors of student learning gains. Later reports documented that student perceptions were also predictors of classroom observation scores. MET did not use the full battery of Tripod student engagement items, but it did include a few that we employ below for measuring happiness and effort in class and whether the teacher inspires students to have an interest in college.

The Tripod student perceptions of teaching that were used in MET and that are the focus of this chapter are grouped in seven scales that we call the 7Cs framework. Two (Challenge and Control) are what we call measures of "press" and the other five (Care, Confer, Captivate, Clarify, and Consolidate) are measures of "support."

- **Challenge** concerns both effort and rigor. It concerns a teacher's insistence that students should work hard and persist in the face of difficulty, for example, "*My teacher accepts nothing less than our best effort*" and "*My teacher wants us to really understand the material, not just memorize it*."
- **Control** concerns the degree to which the class is both well-behaved, for example, "*Students in this class behave the way my teacher wants them to*" and on task, "*Our class stays busy and doesn't waste time*." The connotation is not that teachers are controlling in the sense that they squash student autonomy and expression, but rather in the sense that they are able to manage the class in a way that teaching and learning occur efficiently, without being derailed by misbehavior or distractions.
- **Care** concerns whether the teacher develops supportive relationships with students and is attentive to their feelings. For example, "*My teacher in this class really tries to understand how students feel about things*" or "*My teacher seems to know if something is bothering me*." The Tripod 7Cs conception of care is focused on emotional support. An alternative conception of caring concerns a teacher's commitment to make sure that students succeed. That alternative is not captured by *Care* on its own, but rather by all of the components collectively, especially *Challenge*.
- **Confer** concerns the degree to which the teacher elicits ideas from students and welcomes their feedback. One example is "*My teacher welcomes my ideas and suggestions*." Another is "*My teacher wants us to share our thoughts*." Classrooms that students rate high on *Confer* are more "student centered" than those where only the teacher's perspective is valued.
- **Captivate** pertains to how effectively the teacher stimulates students to be interested in their lessons. A reverse coded item in this category is "*This*

class does not keep my attention—I get bored." A positively worded item is "*My teacher makes lessons interesting*." Items are geared to measure whether the teacher is able to hold the students' attention in class and provide the basis for continuing interest.

- **Clarify** concerns how effectively the teacher is able to help students understand what she is trying to teach them, especially with regard to concepts that students may find difficult to understand. This includes having clear explanations, "*My teacher explains difficult things clearly*," multiple explanations, "*My teacher has several good ways to explain each topic that we cover in this class*," and a commitment to persist until understanding is achieved, "*If you don't understand something, my teacher explains it another way*."
- **Consolidate** concerns making learning coherent, for example, *"My teacher takes time to summarize what we learn each day*," giving feedback, "*The comments that I get on my work in this class help me understand how to improve*," and checking for understanding, "*My teacher checks to make sure we understand what s/he is teaching us."* Hence, *Consolidate* is closely related conceptually to both *Clarify* and *Challenge*.

Each of the 7Cs components is measured by multiple items in the Tripod student survey. MET used one version of the survey for grades 4 and 5 and another for grades 6 and higher. Both versions cover the same 7Cs concepts, although some items are worded more simply for the elementary school version.

Generally, both the FfT and Tripod 7Cs frameworks have components pertaining primarily to communication about rules for time use, procedures, effort, and personal conduct. In addition, both have components pertaining primarily to communication and aspects of instruction associated very directly with implementing the curriculum. For *domains* and their respective *components*, see the Framework Map in Exhibit 4.1 that follows.

PAST LITERATURE

The two-way distinction between Classroom Environment (FfT) or Press (Tripod 7Cs), on the one hand, and Instruction (FfT) or Support (Tripod 7Cs), on the other hand, is reflected in a long tradition of thought on teaching, parenting, and other types of hierarchal relations. Essentially, the first half of the distinction is concerned with the power relations between adults and the students that they teach, supervise, or parent. The other half of the distinction is focused on helping students to achieve understanding, to feel emotionally secure, and to find satisfaction in learning.

In the parenting literature, Diana Baumrind (1966, 1996) contrasts different parenting styles by the degree to which they are "demanding" (related to power) and "responsive" (focused on warmth and various forms of support). She originated the following well known typology in the 1960s: "*Authoritative* parents are both highly demanding and highly responsive, by contrast with *authoritarian* parents, who are highly demanding but not responsive; *permissive* parents, who are responsive but not demanding; and *unengaged* parents, who are neither demanding nor responsive" (1996, p. 412). She writes, "It may be said that the two intertwined generic positive childrearing goals are to foster moral character and optimal competence" (op. cit.).

EXHIBIT 4.1. Framework Map

Framework for Teaching	*TRIPOD 7Cs*

Domain: Classroom Environment
• Creating an environment of respect and rapport
• Managing classroom procedures
• Managing student behavior
	Domain: Press
	• Challenge
	  ◦ Press for Rigor
	  ◦ Press for Persistence
	• Control
	  ◦ Minimize Misbehavior
	  ◦ Promote Good Behavior
	  ◦ Class Stays Busy and on Task

Domain: Instruction • Using questioning and discussion techniques • Using assessment in instruction • Engaging students in learning • Communicating with students • Establishing a culture for learning	**Domain: Support** • Confer • Captive • Clarify • Consolidate • Care

Similarly, scholars who focus on school and classroom environments have distinguished “academic press” from what some call “social support” (Lee & Smith, 1999; Lee, Smith, Perry, & Smylie, 1999), what others call “sense of community” (Shouse, 1996) and still others call personalization (Klem & Connell, 2004). Lee, Smith, Perry, and Smylie (1999) trace the distinction to writers in the early 20th century, including Flexner and Bachman (1918) writing about schools in Gary, Indiana. Baumrind (1996) reaches further back, reminding readers about debates concerning childhood self-determination that we associate with philosophers Hobbs, Rousseau, and Hegel. The issue is the need (and the right) for adults to exert control in order to foster conditions under which they can teach what the child needs to know in order to learn and mature in accordance with societal norms or survival requirements. Lee and colleagues (1999) write, “This report challenges ‘either-or’ proposals for school reform that view academic focus and rigor and social support for students as contradictory strategies. It argues that, to succeed in schools that press them hard to learn, students need strong social support. Conversely, even in the presence of strong social support, students will not learn much unless schools press them to achieve academically” (p. 2).

What do Lee and her colleagues mean by “social support”? Especially because the literature is sometimes unclear, we find it important to emphasize the conceptual distinction between *social or relational* supports, on the one hand, versus *pedagogic* supports and practices, on the other hand. Clearly, we in this chapter are not considering social supports outside the classroom. However, it could be argued that *Care* and *Confer* in the 7Cs framework and perhaps *Engaging Students in Learning*, *Communicating with Students*, and *Establishing a Culture for Learning* in the FfT framework entail social supports. More narrowly pedagogic in nature are *Using Questioning and Discussion Techniques* and *Using Assessment in Instruction* in the FfT framework and *Clarify*, *Captivate*, and *Consolidate* in the Tripod 7Cs framework. The literature is inconsistent on whether the least relational pedagogic practices belong in the support category rather than in the press category or whether they should be included at all in the support-press dichotomy. Whatever reasons there might be for inconsistency in the literature, our conception of support is primarily about instructional supports, some of which are more social or relational than others. Our conception includes all components in the *Instruction* and *Support* domains of the two frameworks.

Now, before moving to a discussion of findings, let us introduce key features of the data and methods that we use.

DATA AND PRIMARY METHODS

All of the data in the chapter are from the 2009–2010 school year of the MET project. Like standard value-added measures, all of the FfT and Tripod 7Cs variables are adjusted to remove variation associated with available measures of student background, including racial and ethnic backgrounds and free lunch status. The reason for the adjustments is to isolate and retain as best we can the variation in the data that is due to teaching and not predictable on the basis of student background characteristics. In addition, all of the data are classroom-level averages, where each classroom supplies one data point for each component of each measure. For example, there were 1,892 classrooms that had data for both FfT and the 7Cs measures during the 2009–2010 school year from which we drew the data for this analysis. The majority of teachers contributed two classrooms, and most classrooms have ratings from two separate FfT observations. Most of the analyses here combine the data for grades 4 through 8 for English and math classes. In addition, we combine value added from state tests and other more cognitively demanding tests that MET used to form a single value-added measure for each classroom.

The primary methods that we use are tabulations and multiple regression analyses. The presentation is organized in such a way that lack of familiarity with multiple regressions should not prevent the reader from gaining a basic understanding. We use simple indicator variables to account for differences associated with grade-level and subject differences. In addition, most regressions are structured to focus on differences between teachers who are colleagues within schools (they include an intercept for each school and they adjust for clustering by teacher). For each FfT measure, the rating that we use for any given classroom is the average for that measure across the multiple times that the classroom was observed. Similarly, each Tripod 7Cs rating is an average from all of the students who responded to the survey in that particular classroom. Finally, unless otherwise indicated, all variables are scaled to have a mean of 0 and a standard deviation of 1 defined on the classroom-level distribution.

DO ADULTS (FfT) AND STUDENTS (TRIPOD 7Cs) AGREE ABOUT TEACHING?

Now that we have introduced the frameworks and connected key concepts to past literature, this section considers how the two frameworks are conceptually and empirically related to one another. We explore how strongly adults using the FfT components agree with students using the Tripod 7Cs. The answers have bearing on how the two might be used together as instructional quality measures.

Conceptually Matching the Frameworks

Please see Exhibit 4.A.1 for the results of a matching exercise. It uses the wording from above that briefly describes each component from each framework. Using arrows, it matches each component to one or more components from the other framework. Matches are based only on the conceptual content of the measures, without reference to the data. They indicate what we regard as the strongest conceptual parallels between the two frameworks.

We use Exhibit 4.2 here in the body of the chapter to summarize the linkages from Exhibit 4.A.1. The left-hand side of Exhibit 4.2 uses the FfT components as headings, while the right-hand side uses the Tripod 7Cs components as headings. For example, on the left-hand side, major heading *Establishing a Culture for Learning* from the FfT is associated with *Challenge* and *Confer* from the 7Cs framework as subheadings, while on the right-hand side, heading *Confer* is conceptually related to four FfT components as subheadings: *Establishing a Culture for Learning, Managing Classroom Procedures, Communicating with Students, and Using Questioning and Discussion Techniques*. The exhibit shows that each component in each framework is related conceptually to one or more components from the other framework. Based on these patterns, we conclude that the frameworks are well matched conceptually, and that the two can form the basis of a coherent discourse on instructional quality.

EXHIBIT 4.2. Cross-Walking the FFT and the 7Cs: Significance Indicators from Multiple Regressions

Predicting FfT Components

- Creating an environment of respect and rapport
  - Care+
  - Control***
- Establishing a culture for learning
  - Challenge***
  - Confer***
- Managing classroom procedures
  - Control***
  - Confer (n.s.)
- Managing student behavior
  - Control***
- Communicating with students
  - Confer*
  - Clarify*
- Using questioning and discussion techniques
  - Challenge**
  - Confer**
  - Consolidate (n.s.)
- Engaging students in learning
  - Challenge*
  - Captivate***
- Using assessment in instruction
  - Clarify***
  - Consolidate (n.s.)

Predicting 7Cs Components

- Care
  - Creating an environment of respect and rapport***
- Confer
  - Establishing a culture for learning**
  - Managing classroom procedures (n.s.)
  - Communicating with students*
  - Using questioning and discussion techniques*
- Captivate
  - Engaging students in learning***
- Clarify
  - Communicating with students***
  - Using assessment in instruction***
- Consolidate
  - Using questioning and discussion techniques*
  - Using assessment in instruction***
- Challenge
  - Establishing a culture for learning***
  - Using questioning and discussion techniques+
  - Engaging students in learning*
- Control
  - Creating an environment of respect and rapport**
  - Managing classroom procedures (n.s.)
  - Managing student behavior***

Note: Two-tailed significance indicators: + 0.10; * 0.05; ** 0.01; *** 0.001.

Empirical Matching

The fact that the frameworks are compatible conceptually does not mean necessarily that data collected using the frameworks will tell the same stories empirically. For example, there could be systematic differences in what students and adults perceive concerning any given issue. To explore the question empirically, we conducted multiple regression analyses of the patterns in Exhibit 4.2.

First, we used each of the FfT components on the left side of Exhibit 4.2 as the dependent variable in a regression equation where the 7Cs components listed under it served as the predictors. Each regression also included a school-level intercept term (a school *fixed effect*) and indicator variables for grade levels and subjects. In addition, as indicated above, each FfT and 7Cs component was adjusted for student background characteristics and scaled to have a mean of 0 and standard deviation of 1. Hence, when predicting a particular FfT component, the estimated coefficients on 7Cs components indicate how strongly each predicts that FfT component, holding constant grade level, subject, and school. Analyses for the right side of the exhibit had the same basic structure as for the left, except that the 7Cs components are the dependent variables and the FfT components are the predictors. Instead of showing regression tables here in the body of the chapter, we simply indicate the two-tailed statistical significance levels on Exhibit 4.2 using symbols that range from "n.s." (for "not significant") to "+" for 90 percent confidence, "*" for 95 percent confidence, "**" for 99 percent confidence, and "***" for 99.9 percent confidence or better. All have the expected signs. (Regression tables are in the Appendix.)

The vast majority of the relationships on Exhibit 4.2 are statistically significant. The main conclusion is that the relationships we expected based simply on our interpretations of the two frameworks are by and large affirmed by patterns in the data. Of course, the full panoply of relationships between all of the FfT components and all of the 7Cs components is much more complicated than we can fully explore. A simple correlation analysis shows that all of the 7Cs components are statistically significantly correlated with all of the FfT components at confidence levels of 95 percent and higher. FfT-to-7Cs correlations range from a low of 0.088 for the relationship between *Consolidate* and *Managing Classroom Procedures* to a high of 0.331 for the correlation between *Control* and *Managing Student Behavior*. These high and low pairings make sense; *Consolidate* and *Managing Classroom Procedures* are probably the least related conceptually of all the components in the two frameworks, and they have the lowest correlation. Conversely, *Control* from the Tripod 7Cs and *Managing Student Behavior* from the FfT both concern student behavior. They are conceptually the most related. Generally, FfT and 7C ratings that are the most conceptually similar tend to be the most highly correlated.

The Special Case of Very Unruly Classrooms

Imagine a classroom in which students are frequently off task and misbehavior appears normal. FfT observers of classrooms that students rated in the bottom quintile on *Control* probably saw such classrooms. When students rated classrooms in the bottom quintile on control, adult observers (none of whom had seen the student ratings) tended to rate it low not only on *Managing Student Behavior*; they tended to rate it low as well on *all* of the FfT components. In fact, classrooms in the bottom quintile of *Control* have such a negative pull on FfT ratings that, when classes in the bottom quintile of *Control* are included in multiple regressions using the full data set, *Control* dominates consistently as the strongest 7Cs predictor for *all* FfT components. Even when multiple regressions omit classrooms rated in the bottom *Control* quintile, where behavior is worst, *Control* is still a strong predictor of FfT ratings.

Figure 4.1 shows findings from eight multiple regressions that omitted the bottom *Control* quintile. FfT components were the dependent variables. The three 7Cs predictors were *Challenge*, *Control*, and a composite of the five components that make up *Support*. The composite for *Support* is the strongest predictor of *Engaging Students in Learning*, *Communicating with Students* and *Using Questioning and Discussion Techniques*, while *Control* is strongest for the other five. If we had included classrooms from the bottom quintile of *Control* in the analysis, *Control* would have been the strongest predictor for all eight FfT components, including those representing the *Instruction* domain in the FfT framework.

To summarize, so far in the chapter we have defined the two frameworks. We find that components of each tend to be more (less) correlated empirically with components from the other that are more (less) conceptually similar. The main exception to this generalization is that *Control* from the 7Cs framework tends to be highly predictive of FfT ratings overall, especially when classrooms from the bottom quintile on *Control* are included in the analysis. Below, we show that *Control* is the strongest predictor of value added as well—stronger than any other component of *either* framework. Similarly,

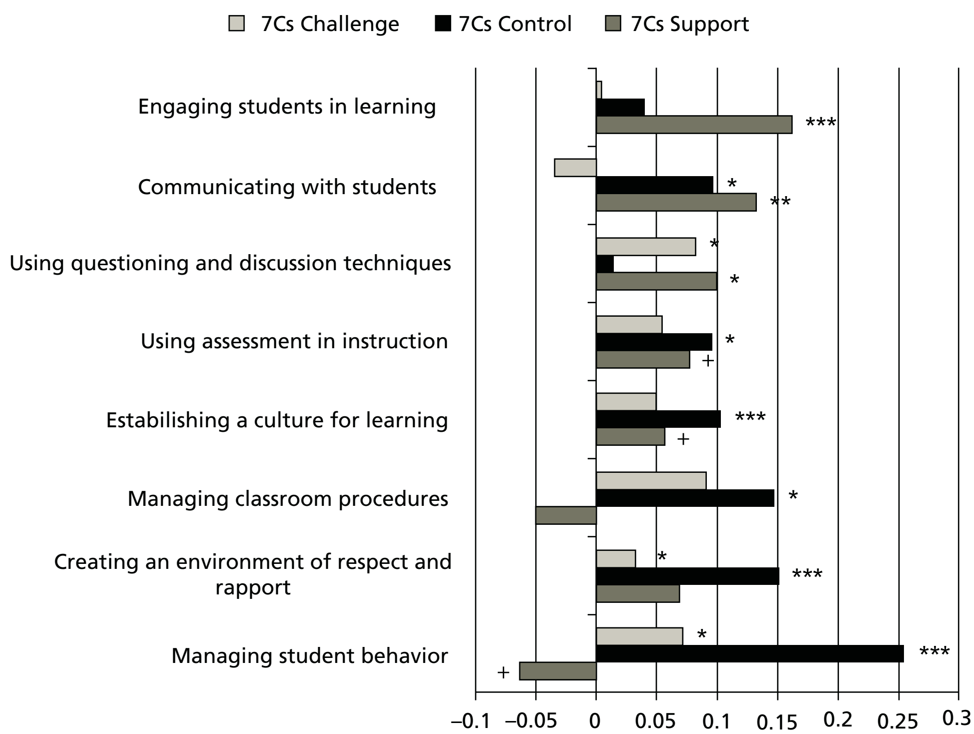

FIGURE 4.1. *Multiple Regression Coefficients Predicting FfT Components Using 7Cs Control, Challenge, and Support*

Note: This figure is computed without classrooms that were in the bottom quintile on Control. The corresponding table in the Appendix shows the results both with and without the bottom quintile on Control in the equations.

Two-tailed significance indicators: + 0.10; * 0.05; ** 0.01; *** 0.001.

Managing Student Behavior is the strongest value-added predictor from the FfT framework. The bottom line is that, when classrooms are out of control and off task, learning is difficult for students and distinguishing clearly among the multiple components of instruction may be almost impossible for adult observers as they try to rate teaching.

PREDICTING VALUE ADDED, HAPPINESS, EFFORT, AND INSPIRATION

MET data include four variables that we consider outcomes of teaching quality: value-added achievement gains, happiness in class, effort in class, and an increased inspiration to attend college. We consider each in turn.

Value Added

Value-added scores for this analysis come from the MET project. They are scaled to have a mean of 0 and standard deviation of 1 defined on the classroom-level distribution for each test by district, subject, and grade. In addition to scores from state accountability exams for math and English language arts (ELA), MET data include value-added scores for the Balanced Assessment in Math (BAM) and the SAT9. The BAM and SAT9 were included in MET to test whether results would differ for accountability and non-accountability exams. In addition, the BAM and SAT9 exams were considered more challenging than most state accountability exams. The overall finding from MET was that patterns were quite similar for accountability and non-accountability exams.

After finding few clear and statistically significant differences between subjects, exams, or grade levels in our work for this analysis, we chose to work with a value-added composite in order to simplify the presentation. First, for math, a composite score for each classroom was set equal to the average of value added from the state math test and the BAM. Then, for ELA, a composite was set equal to the average of the state test and the SAT9. For both math and ELA, we then rescaled to set the composite mean to 0 and the standard deviation equal to 1, defined on the classroom-level MET distribution. Our analysis of value added includes grades four through eight for both ELA and math. Regressions include indicator variables for grades and subjects. They also include school-level intercepts and adjust for clustering by teacher.

We begin with simple tabulations. Figure 4.2 illustrates graphically how each FfT and 7Cs component is related to value added. To construct the figure, we began by creating quintiles for each FfT and 7Cs component. For example, the lowest quintile for *Clarify* contains the bottom 20 percent of classrooms as ranked by that component; the second contains the next 20 percent; and so on up to the top quintile, which contains the 20 percent of classrooms ranked highest for that component. We computed the average value-added score for classrooms in each quintile of each component. Then, for each quintile of each component, we computed the difference between value added for classrooms in that quintile, versus value added for classrooms in its bottom quintile.

FfT and 7Cs components are rank ordered in Figure 4.2 by how much value added in the fifth quintile of a component exceeds value added in its bottom quintile. Several things are apparent. First, the largest fifth-versus-first quintile differences in value added for both FfT and 7Cs frameworks are for *Control* from the 7Cs framework and *Managing Student Behavior* from the

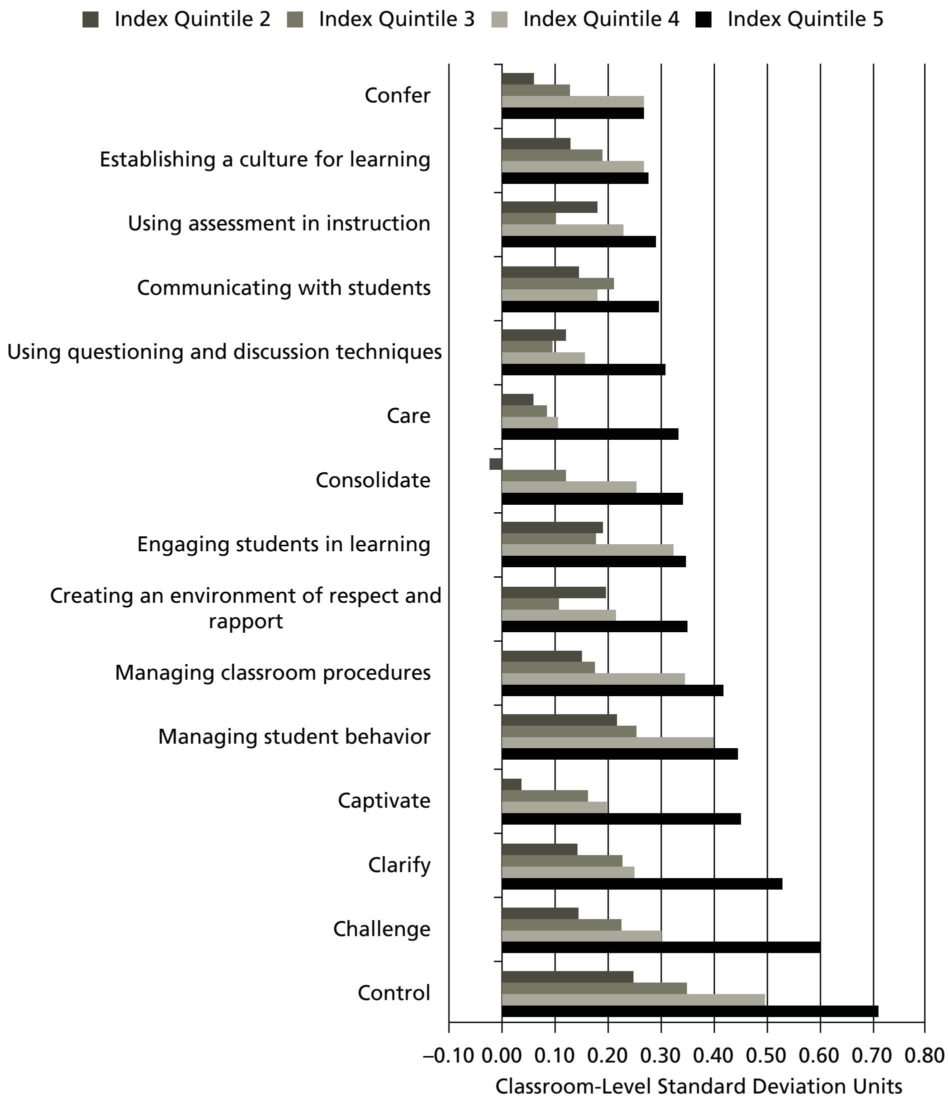

FIGURE 4.2. *Differences in Gains: The Mean Value-Added Test Score Gain for Classrooms in the Second Through Fifth Quintiles of each Respective FfT or 7Cs Component, Minus the Mean Gain in the Bottom Quintile of That Same Component*

FfT framework—the components that measure student behavior management. In addition, for both frameworks, the top few components in the ranking are those related to *Press* or *Classroom Environment*, not *Support* or *Instruction*. Second, while the top four ranked components on the figure are from the 7Cs framework, there is nuance, since the ordering would change if we made a slight change in the ordering criterion. For example, if we ranked using the fourth-versus-first quintile comparisons, *Managing Student Behavior* from the FfT framework would rank second, behind *Control*. The point is that different components distinguish among classrooms in idiosyncratic ways in different parts of their distributions.

Note also that *Control* and *Managing Student Behavior* are the components for which average value added in the second quintile most exceeds that in the first quintile. Finally, notice that value-added levels do not appear to change much when moving from the second to the third quintiles for most components. The greatest differences are in the first and the last steps—from the first to the second quintile, and from either the third to the fourth or the fourth to the fifth, depending on the component. Generally, both FfT and 7Cs components are best at distinguishing value added at their extremes. Even through the middle quintiles, the figure indicates that higher values for 7Cs components are consistently associated with higher value added. The same appears true of *Managing Student Behavior* and *Managing Classroom Procedures*, the FfT components that predict value added most strongly.

Happiness and Effort in Class and College Inspiration

Do FfT and 7Cs components predict *Happiness in Class*, *Effort in Class*, and a teacher who *Inspires Interest in College*? Figure 4.3 shows multiple regression coefficients where the dependent variables are value added (Panel A); "Happy in Class" (Panel B); "Effort in Class" (Panel C); and "Teacher Inspires Interest in College" (Panel D). FfT predictor variables are *Instruction* and *Class Environment* for regressions reported on the left side of the figure. The 7Cs predictor variables are *Support*, *Control*, and *Challenge* in the regressions reported on the right side of the figure.

Consistent with Figure 4.2, Figure 4.3 shows for value added that the FfT composite for *Class Environment* (on the left side) and the 7Cs components *Control* and *Challenge* (on the right side) are stronger predictors of value added than the composites for *FfT Instruction* or *7Cs Support*. Recall that a feature of multiple regressions is that the estimated coefficient for each predictor variable indicates the effect on the dependent variable of changing that particular predictor while the other predictors are held constant. Accordingly, an interpretation of the 7Cs result for value added in Figure 4.3 is, that when holding *Control* and *Challenge* constant, the predicted effect of increasing the

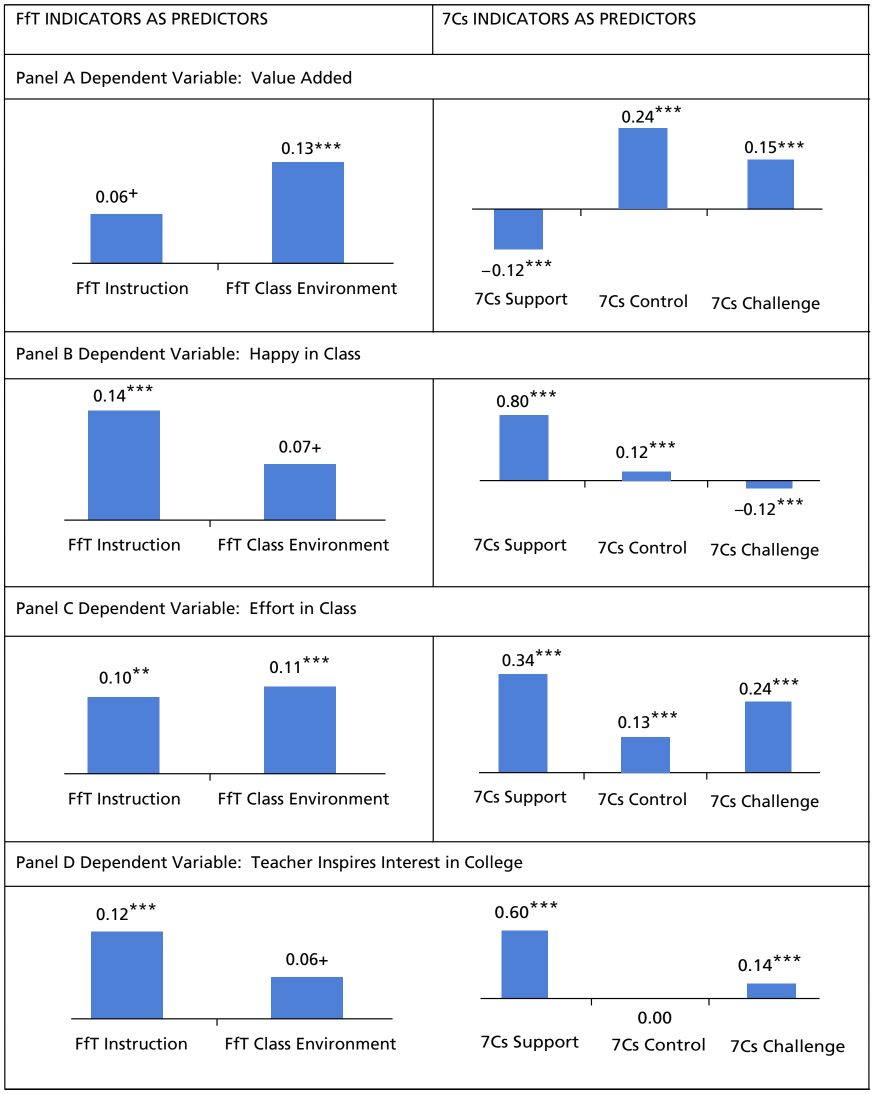

FIGURE 4.3. *Eight Multiple Regressions Predicting Student Outcomes with FfT Domains (Left Side) or 7Cs Domains (Right Side) as Predictors*

Note: Two-tailed significance indicators: + 0.10; * 0.05; ** 0.01; *** 0.001.

composite for *Support* is actually to reduce value added by a modest but statistically significant amount. We speculate below about the reasons.

Panels B, C, and D show a much different pattern. Results for *Happy in Class* in Panel B and *Teacher Inspires Interest in College* in Panel D indicate that *Instruction* and *Support* composites are much stronger predictors of these two outcomes, compared to *Class Environment*, *Control*, and *Challenge*. For *Effort in Class*, Panel C shows that the predictors are similar in their estimated impacts; coefficients for *Instruction* and *Class Environment* are almost identical. And the coefficient for *Support* is almost equivalent to the sum of the coefficients for *Control* and *Challenge*. In other words, a 1 standard deviation change in the *Support* composite is predicted to increase *Effort in Class* by about the same amount as 1 standard deviation increases in both *Control* and *Challenge*.

Another way of contrasting the relationship of value added to teacher quality, versus the relationship of happiness to teacher quality is presented

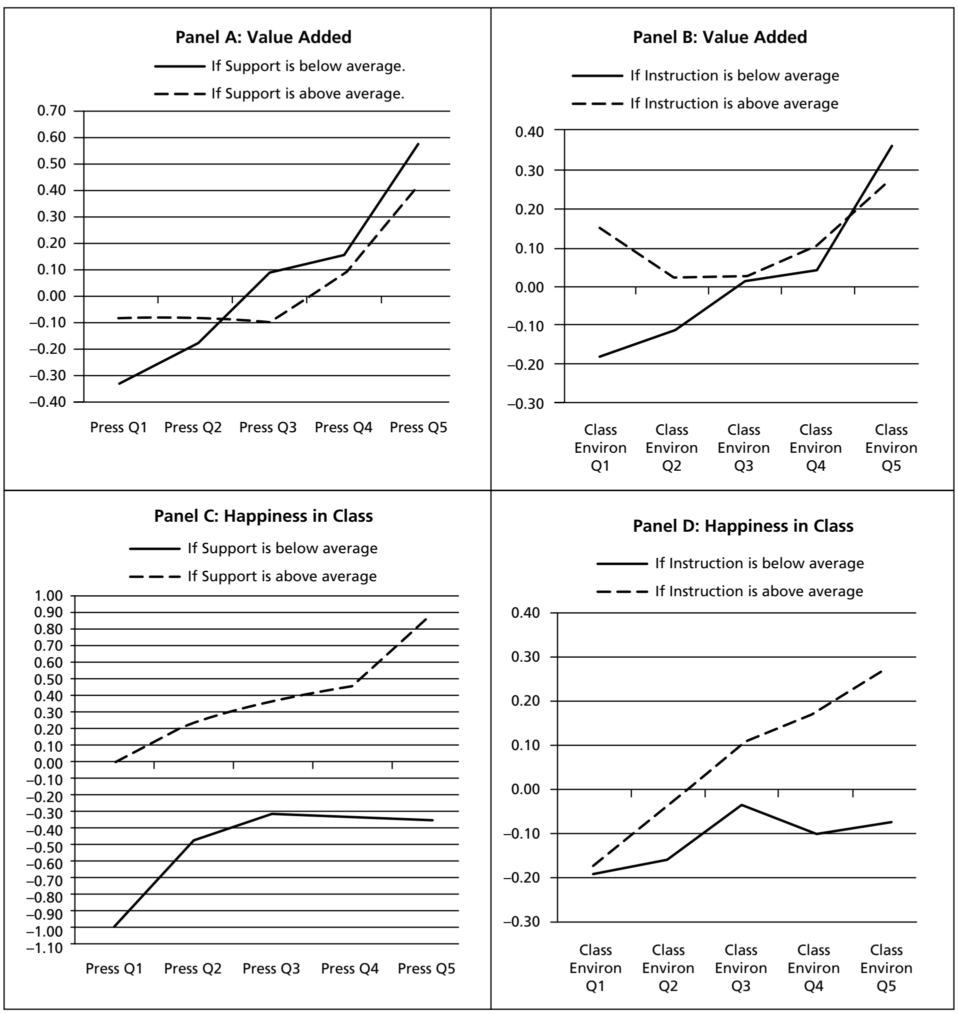

FIGURE 4.4. *Graphs of Actual Value Added and Happiness in Class*

in Figure 4.4. Panels A and C of Figure 4.4 use data from 7Cs metrics, while Panels B and D use data from FfT measures. *Press* in this exhibit is the average of *Control* and *Challenge*. The horizontal axes in Panels A and C represent quintiles for *Press*. Similarly, the horizontal axes for Panels B and D represent quintiles for *Class Environment*. The vertical axes for Panels A and B represent value added, while those for Panels C and D represent *Happiness in Class*.

The dotted lines show relationships between each outcome and *Press* (or *Class Environment*) when *Support* (or *Instruction*) is above average. Conversely, the solid lines show the relationships when *Support* (or *Instruction*) is below average. Hence, moving vertically from the dotted line to the solid line represents a reduction in *Support* or *Instruction* (i.e., from above average to below average), while moving vertically from the solid line to the dotted line represents the opposite.

Note that when *Press* is relatively low, in its first or second quintile, the dotted line is above the solid line. In other words, among teachers who rate relatively low on *Press*, those rated above average on *Support* have higher value added. Indeed, a similar pattern appears in Panel B of the figure, using the FfT framework. However, when *Press* is above average, value added tends to be slightly higher when *Support* is below average. When *Class Environment* is above average (along the horizontal axis of Panel B), the pattern for the solid versus dotted line is mixed, but generally, whether *Instruction* is rated above or below average appears not to matter very much.

The reasons for these patterns are impossible to infer from these data with any certainty. Nonetheless, it is easy to *imagine* an explanation that is relatively simple. Specifically, if a teacher that students rate low on *Press* or that observers rate low on *Class Environment* becomes much more supportive or much better at instruction, students might become more focused and learn more, even if that teacher remains a poor manager of student behavior. However, if a teacher is already quite challenging and the class is almost always well behaved and on task, then becoming more supportive might actually lower the sense of urgency in the class and some students might relax and learn less. Again, while thought provoking, any explanation at this point remains speculation, since there is no way to know from the available data.

Based on the preceding, one might be tempted to conclude that being high on *Press* and below average on *Support* is a good thing. Or, referencing Panel B of Figure 4.4, one might conclude that as long as *Class Environment* is average or above, *Instruction* is rather unimportant.

But not so fast! Before making so rash a judgment, see Panels C and D. They indicate students are happier in classrooms where *Support* as rated by the 7Cs and *Instruction* as rated by FfT metrics are above average. An analogous exhibit for *Teacher Inspires Interest in College* would appear quite similar. It seems quite reasonable to expect that students who spend lots of time in supportive classrooms will grow to love learning more and be more prone to become lifelong, voluntary learners. Prioritizing *Press* and *Classroom Environment* in order to maximize value added, while neglecting to improve *Support* and *Instruction*, would surely be a short-sighted strategy. There needs to be balance.

It turns out that the relative strength of academic *Press* metrics as predictors of learning is consistent with findings in past research. For example, Lee and Smith (1999, p. 907) report: "we found that, on average, social support is positively but modestly related to learning. However, both learning and the relationship between social support and learning are contingent on the academic press of the school students attend." Similarly, Shouse (1996, p. 47) reports: "for most schools, academic press serves as a key prerequisite for the positive achievement effects of communality." Definitions and measures of support and "communality" in past studies pertain more to the relational aspects of support than to the instructional aspects. Still, across studies, there appears to be consistency in the finding that *Press* tends to be the stronger predictor of learning.

It seems highly plausible, indeed likely, that sustaining high *Press* (*Control* and *Challenge*) without intimidation and coercion requires providing a significant degree of *Support* (*Care*, *Confer*, *Captivate*, *Clarify*, and *Consolidate*). Similarly, achieving a high-quality *Classroom Environment* without relying on fear probably requires a relatively high quality of *Instruction*. To establish these propositions definitively would require longitudinal data, generated experimentally. Still, we can use the data that we have to deepen our intuition.

Achieving Order without Intimidation and Coercion

In order to consider whether *Support* might provide a foundation for *Press*, we ask, "Which components of the *Instruction* domain are the strongest predictors of *Classroom Environment* in the FfT framework?" and "Which components of the *Support* domain are the strongest predictors of *Press* in the Tripod 7Cs framework?" The question is whether patterns are consistent with the hypothesis that the quality of *Support* and *Instruction* might influence value added (and other outcomes) indirectly by affecting student behavior and focus in the classroom, as measured by the components of *Classroom Environment* and *Press*.

We conducted multivariate regressions to help us judge the plausibility of this hypothesis. Figure 4.5 shows the results. First, it is interesting to note that the pattern for predicting *Control* is clearly different from that for predicting *Challenge*. For *Control*, the three strongest predictors are *Clarify*, *Confer*, and *Captivate*, with *Clarify* as the strongest. A likely interpretation is that explaining concepts clearly (*Clarify*), talking with students (*Confer*) and making lessons interesting (*Captivate*) helps keep the class orderly and on task. For *Challenge*, *Captivate* is not a predictor at all. Apparently, making lessons interesting is not how teachers press students to think rigorously and persist in the face of difficulty. Instead, *Challenge* is almost equally (and quite strongly) predicted by *Clarify* and *Consolidate*. Both of the latter pertain to ways of helping students achieve understanding—explaining material clearly, summarizing, and checking for understanding. *Confer* predicts *Challenge* as well, but only half as strongly as *Clarify* and *Consolidate* and about equally as strongly as it (i.e., *Confer*) predicts *Control*. For both *Challenge* and *Control*, *Care* enters the multiple regression with a small negative and statistically significant coefficient. The concept of *Care* in the 7Cs framework is closely related to emotional support. So the finding that, other things being equal, more *Care* predicts slightly less *Control* and *Challenge* is not really surprising. Indeed, it reminds us of the finding in Figure 4.3, Panel A, concerning the negative role of *Support* in predicting value added when holding *Control* and *Challenge* constant.

For the FfT part of the analysis, our predictors are components from the *Instruction* domain whose labels constitute vivid *action* statements of what the teacher is doing instructionally—i.e., *Communicating with Students*, *Using Assessment in Instruction*, and *Using Questioning and Discussion Techniques*.

The dependent variables are the other five FfT components. Again, just as above, the question is whether predictive patterns are consistent with the hypothesis that components from the *Instruction* domain are affecting conditions measured by components from the *Classroom Environment* domain. In the same spirit, we examine how the three *Instruction* components that describe actions in their titles predict the two that have “learning” in their titles. The question is whether higher ratings on the instructional action components predict FfT rater perceptions that there is a culture of learning and that students are intellectually engaged.

Figure 4.5 indicates that all three of the FfT components with actions in their titles make distinct contributions to predicting all five of the other FfT

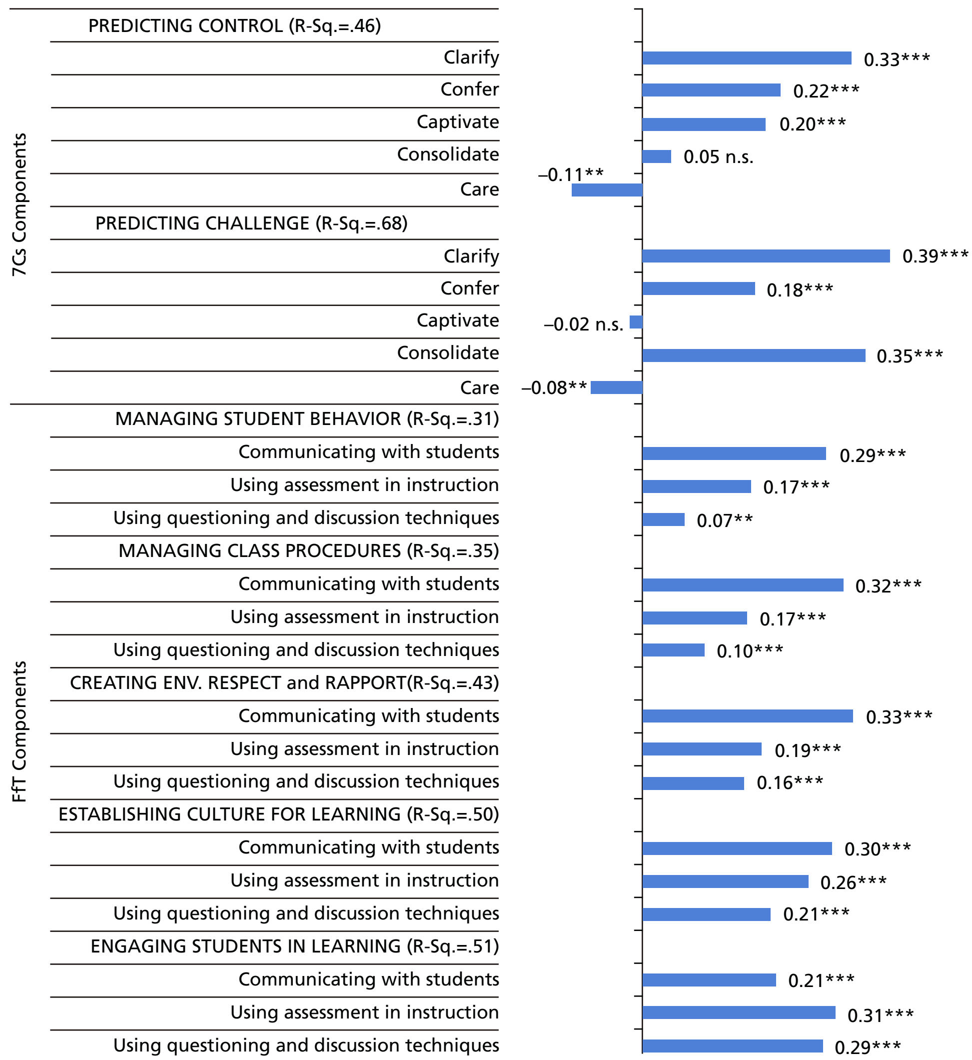

FIGURE 4.5. *Predicting Control, Challenge, and Classroom Environment Components Using Support and Instructing Components*

Note: Two-tailed significance indicators: + 0.10; * 0.05; ** 0.01; *** 0.001.

measures. For each of the standard four *Classroom Environment* components (treating *Establishing a Culture for Learning* as the fourth), *Communicating with Students* is the strongest predictor. Importantly for our hypothesis, among the FfT *Instruction* components, *Communicating with Students* is most associated conceptually with making success feasible for students who might otherwise struggle. It concerns "clarity of lesson purpose; clear directions and procedures specific to the lesson activities; absence of content errors and clear explanations of concepts and strategies; and correct and imaginative use of language." Hence, it closely parallels *Clarify* in the 7Cs framework.

Similar to findings for the 7Cs above, our interpretation is that high-quality delivery of instruction is surely among the practices that enable teachers to achieve orderly and focused classroom environments without needing to use intimidation and coercion.

To summarize, an important finding in this chapter is the strong roles of *Classroom Environment*, *Control*, and *Challenge* in predicting value added. The fact that *Support* and *Instruction* components help predict these measures in the manner shown in Figure 4.6 challenges any presumption that *Control* and *Managing Student Behavior* are best achieved through heavy-handed, coercive methods. Instead, it appears likely that both *Control* and *Managing Student Behavior* are best achieved—and indeed most effective—when students and observers alike perceive clarity in the delivery of the instruction and free-flowing communication with students who might otherwise struggle, misbehave, and go off task. It could be that improving teachers' content knowledge and associated methods for helping students with difficult material could be the most effective ways of maintaining the types of orderly, on-task classrooms that produce the most learning.

MAINTAINING DATA QUALITY

Districts around the nation are becoming more serious about using data to inform their efforts toward improvement. But data will lose their validity and value if too many of those who rate teaching fail to take care or try to manipulate outcomes. Especially in a context where stakes are high, vigilance to maintain data quality is important. Since there were no stakes for teachers or students in the MET project, there was no motive to systematically distort responses. However, the need to monitor for biases and distortions will rise as more school systems use these measures in their accountability formulas. Students in a school that uses student surveys in teacher evaluation formulas might try to influence evaluation outcomes by either inflating or deflating their responses. Administrators or coaches who rate teachers might try to low-ball ratings for teachers they want to dismiss. Or, for teachers they think deserve the benefit of the doubt, they might exaggerate ratings in order to offset low ratings from other sources, such as student ratings or value-added scores.

In either case, officials can monitor for patterns of inconsistency between 7Cs and FfT ratings in order to detect when greater scrutiny is warranted.

When 7Cs composites and FfT ratings are scaled to have standard deviations of 1 and means of 0, the difference between the standardized 7Cs and FfT composites should on average be 0, with a bell-shaped distribution around that average. Typical differences can be benchmarked using large data sets such as the MET data or a combined Tripod and FfT database for schools using both assessment systems. Irregularities can be fairly easy to detect.

Imagine measuring 7Cs-versus-FfT disagreement at the classroom level by subtracting the standardized 7Cs composite rating from the standardized FfT composite. We implemented this procedure for the MET data to examine the patterns. Column A in Table 4.1 represents the likelihood in the MET

TABLE 4.1. Detecting Irregularities: Probabilities That the Difference Between FfT and 7Cs Composites Would Fall Repeatedly in Selected Ranges by Chance; Column A Is Based on MET Data

	Number of Repetitions in Same Range				
	A Single Occurrence	Two	Three	Four	Five
Column	A	B	C	D	E
Difference Between FfT and 7Cs Composites	Pattern probabilities if each rating is independent				
0 or higher	0.5000	0.2500	0.1250	0.0625	0.0313
Over 0.10 standard deviation	0.4500	0.2025	0.0911	0.0410	0.0185
Over 0.25 standard deviation	0.4000	0.1600	0.0640	0.0256	0.0102
Over 0.5 standard deviation	0.3100	0.0961	0.0298	0.0092	0.0029
Over plus 1 standard deviation	0.1700	0.0289	0.0049	0.0008	0.0001
Between plus and minus 0.25 standard deviation	0.1500	0.0225	0.0034	0.0005	0.0001
Over 2.0 standard deviations	0.0300	0.0009	0.0000	0.0000	0.0000

data of observing a difference in each range listed on the left of the table. For example, the range “0 or higher” means that the FfT rating was higher than the 7Cs rating. The likelihood of this happening in the MET data is almost exactly 0.50—the same as flipping a coin. The likelihood that the FfT rating exceeds the 7Cs rating by 0.10 standard deviation is 0.45. Skipping a few lines down, we see that the likelihood of more than 1 standard deviation is 0.17 and for more than 2 standard deviations the number is 0.03. In addition, we consider the likelihood of falling in a rather narrow range around the mean: between plus and minus 0.25 of a standard deviation. The likelihood of this happening is only 0.15.

The other columns of Table 4.1 show the probabilities that multiple independent classrooms would have differences in the same range. Imagine, for example, that an administrator works in a school that surveyed students. The same administrator rates teachers. If the administrator has two classrooms to judge and is perfectly even-handed—in other words, she judges each completely on its own terms—and the students give an honest appraisal as well, then the likelihood that her FfT rating will be higher than the students’ 7Cs rating for both teachers is 0.25. In other words, it will happen about a quarter of the time. However, if it keeps happening, there is likely to be some irregularity. The likelihood that her FfT rating will be higher than the students’ 7Cs rating five times in a row if there are no irregularities is only 0.03—only three times in one hundred. Scrutiny would seem warranted.

The likelihood of landing repeatedly in other ranges that Table 4.1 shows is even smaller. It is interesting that excessive matching between FfT and the 7Cs ratings is also an irregularity. Falling within 0.25 standard deviations of agreement five times in a row will happen merely by chance only one in ten thousand times. The laws of probability are quite robust. Especially when there are several classrooms evaluated using both observational and student survey tools, levels of agreement and disagreement between the two methods will follow probabilistic patterns that make systematic irregularities readily detectable.

DISCUSSION AND IMPLICATIONS

This chapter set out to do three things: first, to examine whether specific domains and components of the FfT and Tripod 7Cs frameworks are well-matched not only conceptually, but also empirically; second, to explore which components of both frameworks predict value added and three measures of student engagement that are of concern to parents and educators alike; and third, to suggest how data collected using the FfT and Tripod 7Cs frameworks can be used in combination to monitor the implementation fidelity of both.

Do Adults and Students Agree about Teaching?

We find that the conceptual overlap between the frameworks is substantial and that empirical patterns in the data show similarities in adult and student assessments at the classroom level. Educators can cross-walk the two frameworks to consider the management of student conduct (e.g., *Control* from the 7Cs and *Managing Student Behavior* from the FfT); classroom-level relationships (e.g., *Care* from the 7Cs and *Creating a Climate of Respect and Rapport* from the FfT); delivery of the curriculum (e.g., *Clarify* and *Consolidate* from the 7Cs and *Using Questioning and Discussion Techniques* and *Using Assessment in Instruction* from the FfT); eliciting student perspectives and inviting help-seeking in class (e.g., *Confer* and *Clarify* from the 7Cs and *Communicating with Students* from the FfT); and making learning attractive (e.g., *Captivate* from the 7Cs and *Engaging Students in Learning* from the FfT). More examples could be listed. Generally, the 7Cs *Press* components are related to the FfT *Classroom Environment* components, and the 7Cs *Support* components are related to the FfT *Instruction* components.

Implications Based on our analysis, we judge it quite practical for educators to use both frameworks in their reflections on teaching and to draw data from both measurement systems to identify challenges, assess progress, and set goals.

Predicting Value Added and Engagement

The data that were available for this analysis included indices for value-added achievement gains, happiness in class, effort in class, and the degree to which the teacher inspires an interest in college. We found that all four were predicted in interesting ways by components of both the FfT and 7Cs frameworks.

The chapter augments MET reports by going inside the composite metrics to study how individual components and domains relate to value added. We also pay more attention to the engagement measures than was practical in the MET reports. Each of the eight FfT components and seven 7Cs components is correlated to a statistically significant degree with value added, but some more strongly than others. For both the FfT and the 7Cs frameworks, the component most conceptually associated with student conduct management was the strongest predictor of value added. In addition, student respondents to Tripod surveys and observers using the FfT protocol tend to agree when behavior is a problem. The correlation between FfT and 7Cs metrics for conduct management—*Control* and *Managing Student Behavior*—was the strongest between the frameworks. A notable finding is that being in the bottom quintile for these conduct measures, especially for *Control*, is especially problematic. Aside from the disruptive impacts on teaching and learning, problems with behavior management may restrict an FfT observer's ability to judge instruction along the multiple dimensions that the FfT aims to measure. We found that classrooms rated in the bottom quintile on *Control* were rated low on all FfT components, and to a greater degree students rated the same classrooms low on the 7Cs components.

An unexpected but robust finding is that holding constant *Control* and *Challenge* while increasing *Support* tends to depress value added. Similarly, *Instruction* has a smaller predicted impact on value added than *Classroom Environment*. We speculate that when students rate a teacher low on both *Support* and *Press* or when an observer sees reasons to rate a classroom low on both *Instruction* and *Class Environment*, increasing the level of *Support* or the quality of *Instruction* may encourage or enable students to pay more attention to their studies. However, when instructional quality is already high for most components, increasing *Support* might relieve some of the stress that drives students to focus and persevere. For the time being, this explanation remains speculation but seems consistent with patterns in the data.

While *Control* and *Challenge* from the 7Cs framework and *Classroom Environment* from the FfT were the strongest predictors of value added, components from the *Support* domain of the 7Cs framework and the *Instruction* domain of the FfT framework most strongly predicted *Happiness in Class* and *Teacher Inspires Interest in College. Effort in Class* was predicted by a relatively balanced combination of *Support* and *Press* factors.

Finally, our findings are consistent with the possibility that *Support* and *Instruction*—especially *Clarify* and *Communicating with Students*—might be important enablers for *Press* and *Classroom Environment*.

Implications The finding that *Control*, *Challenge*, and the components of *Classroom Environment* are stronger and more consistently positive predictors of value added than the components of *Support* and *Instruction* may tempt some educators to increase their focus on heavy-handed, coercive ways of managing classrooms. There are two reasons that this may be a mistake. First, most of us, including parents, educators, and others, want classroom climates that foster a love of learning. Findings here concerning predictors of happiness, effort, and interest in college caution us to value classroom experiences associated with higher values in the *Instruction* and *Support* domains. Second, it seems highly likely that better performance in the *Instruction* and *Support* domains might actually buttress performance in the *Press* and *Classroom Environment* domains. Hence, maintaining a balance seems important. If any one or two components should be the focus, *Clarify* from the 7Cs framework and *Communicating with Students* from the FfT framework are prime candidates to consider. This is because they not only help to predict the non-value-added outcomes—happiness, effort, and college inspiration—but they are also strong predictors of *Control*, *Challenge*, and *Classroom Environment*, which in turn, are strong predictors of value added.

Maintaining Data Quality

One hears a good deal of informal conversation these days concerning the ways that measurement tools can be misused, especially under high-stakes conditions. Therefore, it is important to make sure that stakeholders are well prepared to use tools correctly and that carelessness or intentional misuse are discouraged or curtailed. It is difficult to know how severely the incentives entailed with high-stakes use of FfT and Tripod instruments could distort the ways that people use them. Clearly, reports of cheating over the past year on standardized testing in the Atlanta school system should make us cautious.

There are multiple ways to detect irregularities. Using large benchmarking data sets, it is possible to check how frequently disagreement between FfT and 7Cs composites will tend to be in particular ranges. There are some ranges that occur very rarely, even for a single class. For example, in the MET data, the difference between FfT and 7Cs ratings exceeds 2 standard deviations in only 3 percent of classrooms. Such a large difference should trigger at least a modest bit of examination. However, even more important, having multiple classrooms fall consistently within the same range of FfT-7Cs disagreement should be a very rare event. Table 4.1 shows, for example, that finding five classrooms where the FfT rating in standard deviation units exceeds the 7Cs rating all five times should be a rare event. An administrator who consistently rates teachers higher than the students do is very likely out of calibration, and an intervention is almost certainly warranted.

Limitations

Despite the unusual richness of the data, the work that we discuss in this chapter is very much work in progress. Because all of the data are cross-sectional and generated from natural variation, not through planned or experimental variation, estimates must necessarily be interpreted as correlational, not causal. Causal statements in this context are judgments, not findings. In addition, this has been an aggregate analysis. We have lumped together different subjects, different tests, different grades, and different schools and districts. Whether our conclusions will apply to more homogeneous categories of analysis is beyond the scope of this chapter, but nonetheless important to address in future work. Finally, it is important to note that the MET data were collected under special conditions. Raters were trained and monitored to ensure that they scored classrooms correctly. Student surveys were administered according to protocol, and data records were carefully managed. The reliability and validity of the data deepened upon quality control. The findings reported here are unlikely to apply in instances when observation and data collection procedures are not of high quality.

CONCLUSION

This chapter concerns ideas and tools for helping educators to reflect on their work, refine their craft, and increase their effectiveness. We show that concepts and tools from the Framework for Teaching and Tripod survey assessments—tools that the authors of the chapter have designed independently over many years—are quite compatible. Furthermore, MET data collected and organized using the two frameworks show similar empirical relationships to student engagement and learning. Most prominent, analysis of data from both frameworks shows that *classroom management* is the strongest predictor of achievement gains. In addition, for each framework, the teaching component associated with *clarity* is the strongest predictor of effective classroom management.

A central finding is that the challenging and structured teaching practices that most effectively raise test scores are different from the mix of caring and emotionally supportive practices that most effectively foster happiness, voluntary effort, and inspiration to attend college. This does not mean we should downplay the importance of raising test scores. Remember that reading and math scores measure skills for which employers will someday pay and upon which families will someday depend. But schools should strive to achieve a balance between the types of *Press* most strongly associated with short-term growth in measurable skills versus the types of *Support* that foster a healthy and optimistic outlook on life and learning. The ideas and evidence that we examine in this chapter can help us identify and support balanced teaching that fosters the multiple skills and orientations students need to succeed.

Finally, our first priority (and the main reason we do this work) is to help educators to improve instructional quality. However, as the tools become used increasingly for accountability purposes, not simply instructional improvement, the integrity of the data needs to be protected. Monitoring becomes important. We show near the end of the chapter that comparing FfT and Tripod 7Cs ratings from clusters of individual classrooms is a way to discover patterns of irregularity that may warrant careful scrutiny. Data can be standardized in ways that enable analysts to estimate the likelihood that the pattern of FfT and Tripod 7Cs ratings from a particular batch of classrooms could have occurred normally, without systematically inappropriate forces intervening.

Recall that we opened the chapter by asking why some teachers routinely produce more learning than others do. Having reviewed this chapter, what is your answer? What are some implications for your work?

NOTE

1. Domains 2 and 3 each have five components in the FfT framework. MET omitted one from each domain. The omitted components from these domains are "organizing physical space" and "demonstrating flexibility and responsiveness." More about the framework and components, elements and performance levels can be found in *Enhancing Professional Practice* (Danielson, 2007) or at www.danielsongroup.org.

REFERENCES AND ADDITIONAL RESOURCES

Baumrind, D. (1966). Effects of authoritative parental control on child behavior. *Child Development*, *37,* 887–907.

Baumrind, D. (1996, October). The discipline controversy revisited. *Family Relations, 45*(4), 405–414.

Cantrell, S., & Kane, T. J. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study*. Seattle, WA: Bill & Melinda Gates Foundation.

Chart, H. with Kendall-Taylor, N. (2008). *Reform what? Individualist thinking in education: American cultural models on schooling.* Washington, DC: Frameworks Institute.

Danielson, C. (2007). *Enhancing Professional Practice: A framework for teaching*. ASCD.

Danielson, C. (2013). *The Framework for Teaching Evaluation Instrument.* Princeton, NJ: The Danielson Group.

Ferguson, R. F. (2012). Can student surveys measure teaching quality? *Phi Delta Kappan, 94*(3), 24–28.

Flexner, A., & Bachman, F. (1918). *The Gary schools: A general account.* New York, NY: General Education Board.

Ho, A. D., & Kane, T. J. (2012). *The reliability of classroom observations by school personnel*. Seattle, WA: Bill & Melinda Gates Foundation.

Kane, T. J., McCaffrey, D. F., & Staiger, D. O. (2010). *Learning about teaching: Initial findings from the Measures of Effective Teaching Project*. Seattle, WA: Bill & Melinda Gates Foundation.

Kane, T. J., McCaffrey, D. F., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation.

Klem, A. M., & Connell, J. P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health, 74*(7), 262–273.

Lee, V. E., & Smith, J. B. (1999). Social support and achievement for young adolescents in Chicago: The role of school academic press. *American Educational Research Journal, 36*(1), 907–945.

Lee, V. E., Smith, J. B., Perry, T. E., & Smylie, M. A. (1999). *Social support, academic press, and student achievement: A view from the middle grades in Chicago*. Chicago, IL: Consortium on Chicago School Research.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005, March). Teachers, schools and academic achievement. *Econometrica, 72*(2), 417–458.

Shouse, R. C. (1996). Academic press and sense of community: Conflict, congruence, and implications for student achievement. *Social Psychology of Education, 1*(1), 47–68.

APPENDIX

EXHIBIT 4.A.1. Comparing the Framework for Teaching and Tripod 7Cs

Framework for Teaching

Creating an environment of respect and rapport. For example: respectful talk, active listening, and turn taking; acknowledgment of students' backgrounds and lives outside the classroom; body language indicative of warmth and caring; physical proximity; politeness and encouragement; and fairness.

Managing student behavior. For example: clear standards of conduct, possibly posted, and possibly referred to during a lesson; absence of acrimony between teacher and students concerning behavior; teacher awareness of student conduct, including preventive awareness; absence of misbehavior; and reinforcement of positive behavior.

Managing classroom procedures. For example: smooth functioning of all routines; little or no loss of instructional time; students playing an important role in carrying out the routines; students knowing what to do, where to move.

Establishing a culture for learning. For example: belief in the value of what is being learned; high expectations supported through both verbal and nonverbal behaviors, for both learning and participation; expectation of high-quality work on the part of students; expectation and recognition of effort and persistence on the part of students; and high expectations for expression and work products.

Using questioning and discussion techniques. For example: questions of high cognitive challenge, formulated by both students and teacher; questions with multiple correct answers, or multiple approaches,even when there is a single correct response; effective use of student responses and ideas; discussion, with the teacher stepping out of the central, mediating role; focus on the reasoning exhibited by students in discussion, both in give-and-take with the teacher and with their classmates; high levels of student participation in discussion.

Communicating with students. For example: clarity of lesson purpose; clear directions and procedures specific to the lesson activities; absence of content errors and clear explanations of concepts and strategies; and correct and imaginative use of language.

Engaging students in learning. For example: students show enthusiasm, interest, thinking, problem solving, etc.; learning tasks that require high-level student thinking and invite students to explain their thinking; students highly motivated to work on all tasks and persist,even when the tasks are challenging; students actively "working," rather than watching while the teacher "works"; and suitable pacing of the lesson, neither dragged out nor rushed, with time for closure and student reflection.

Using assessment in instruction. For example: the teacher paying close attention to evidence of student understanding; the teacher posing specifically created questions to elicit evidence of student understanding; the teacher circulating to monitor student learning and to offer feedback; and students assessing their own work against established criteria.

Tripod 7Cs

Care concerns whether the teacher develops supportive relationships with students and is attentive to their feelings. For example, "My teacher in this class really tries to understand how students feel about things" or "My teacher seems to know if something is bothering me." The 7Cs conception of care is focused on emotional support. An alternative conception of caring concerns a teacher's commitment to make sure that students succeed. That alternative is captured by all of the Cs, collectively.

Control concerns the degree to which the class is both well-behaved, for example, "Students in this class behave the way my teacher wants them to" and on task, for example, "Our class stays busy and doesn't waste time." The connotation is not that teachers are controlling in the sense of squashing student autonomy and expression, but rather that they are able to manage the class in a way that teaching and learning occur efficiently, without being derailed by misbehavior or distractions.

Confer concerns the degree to which the teacher elicits ideas from students and welcomes their feedback. One example is "My teacher welcomes my ideas and suggestions." Another is "My teacher wants us to share our thoughts." Classrooms that students rate high on Confer are more "student centered" than those where only the teacher's perspective is valued.

Challenge concerns both effort and rigor. It concerns a teacher's insistence that students should work hard and persist in the face of difficulty, for example, "My teacher accepts nothing less than our best effort," and think hard, for example, "My teacher wants us to really understand the material, not just memorize it."

Consolidate concerns making learning coherent, for example, "My teacher takes time to summarize what we learn each day," giving feedback, "The comments that I get on my work in this class help me understand how to improve," and checking for understanding, "My teacher checks to make sure we understand what s/he is teaching us." Hence, Consolidate is closely related conceptually to both Clarify and Challenge.

Captivate pertains to how effectively the teacher stimulates students to be interested in their lessons. A reverse coded item in this category is, "This class does not keep my attention—I get bored." A positively worded item is, "My teacher makes lessons interesting." Items are geared to measure whether the teacher is able to hold the students' attention in class and provide the basis for continuing interest.

Clarify concerns how effectively the teacher is able to help students understand what she is trying to teach them, especially with regard to concepts that students may find difficult to understand. This includes having clear explanations, "My teacher explains difficult concepts clearly," multiple explanations, "My teacher has several good ways to explain each topic that we cover in this class," and a commitment to persist until understanding is achieved, "If you don't understand something, my teacher explains it another way."

TABLE 4.A.1. Backup for Exhibit 4.2, Column 1

	Creating an environment of respect and rapport	Establishing a culture for learning	Managing classroom procedures	Managing student behavior	Communicating with students	Using questioning and discussion techniques	Engaging students in learning	Using assessment in instruction
Column	**1.**	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**
Care	0.051+							
Control	0.255***		0.272***	0.334***				
Challenge		0.115***				0.103**	0.089*	
Confer		0.111***	−0.033		0.096*	0.137**		
Clarify					0.102*			0.174***
Consolidate						−0.047		0.005
Captivate							0.143***	
Subject and Level Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

(*continued*)

(*Table 4.A.1 continued*)

	Creating an environment of respect and rapport	Establishing a culture for learning	Managing classroom procedures	Managing student behavior	Communicating with students	Using questioning and discussion techniques	Engaging students in learning	Using assessment in instruction
Column	**1.**	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**
School intercepts	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.22**	0.133	0.191**	0.101	0.156+	0.122	0.113	0.128
R-Square	0.285	0.260	0.274	0.309	0.216	0.199	0.231	0.225
Adj. R-Square	0.194	0.165	0.181	0.221	0.116	0.096	0.132	0.126

Note: Regressions behind Column 1 of Exhibit 4.2 in the body of the chapter. Each school has a separate intercept term and all regressions have level (i.e., elementary versus middle school) and subject (i.e., ELA versus math) indicator variables.

N = 1,175 teachers; 1,892 classrooms; standard errors adjusted for clustering by teacher.

Two-tailed significance indicators: + 0.10; * 0.05; ** 0.01; *** 0.001.

TABLE 4.A.2. Backup for Exhibit 4.2, Column 2

	Care	Confer	Captivate	Clarify	Consolidate	Challenge	Control
Column	**1.**	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**
Managing classroom procedures	0.18***						0.099**
Communicating with students		0.101**				0.109***	
Establishing a culture for learning		0.017					0.04
Creating an environment of respect and rapport		0.07*		0.109***			
Using questioning and discussion techniques		0.059*			0.068*	0.051++	
Using assessment in instruction			0.195***			0.065*	
Engaging students in learning				0.107***	0.098***		
Managing student behavior							0.238***
Subject and level indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School-specific intercepts	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.249*	0.088	0.167++	0.134	0.248**	0.172++	−0.032
R-Square	0.217	0.258	0.228	0.232	0.247	0.266	0.274
Adj. R-Square	0.118	0.163	0.131	0.135	0.151	0.171	0.274

Note: Regressions behind Column 2 of Exhibit 4.2 in the body of the chapter. Each school has a separate intercept term and all regressions have level (i.e., elementary versus middle school) and subject (i.e., ELA versus math) indicator variables.

N = 1,175 teachers; 1,892 classrooms; standard errors adjusted for clustering by teacher.

Two-tailed significance indicators: + 0.10; * 0.05; ** 0.01; *** 0.001.

TABLE 4.A.3. Regressions Behind Figure 4.1

	Creating an environment of respect and rapport	Managing classroom procedures	Managing student behavior	Establishing a culture for learning	Using questioning and discussion techniques	Communicating with students	Engaging students in learning	Using assessment in instruction
Column	**1.**	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**
Panel A: Regressions exclude classrooms for which Control was in the bottom quintile. This version is shown on Exhibit 4.2.								
Support	0.069+	−0.050	−0.062+	0.057	0.100*	0.132**	0.162***	0.077+
Control	0.151***	0.148***	0.255***	0.103*	0.014	0.098*	0.040	0.096*
Challenge	0.032	0.092*	0.072*	0.050	0.082*	−0.034	0.004	0.054
Constant	0.279***	0.203**	0.155*	0.210*	0.133	0.188+	0.182*	0.138
R-Square	0.257	0.248	0.272	0.258	0.210	0.225	0.232	0.233
Adj. R-Square	0.139	0.129	0.156	0.140	0.084	0.102	0.110	0.111
Classrooms	1554	1554	1554	1554	1554	1554	1554	1554
Teachers	1114	1114	1114	1114	1114	1114	1114	1114
Schools	206	206	206	206	206	206	206	206

Panel B: Regressions below include even the bottom quintile on Control. Not shown in the body of the chapter.

Support	0.014	−0.096*	−0.095**	0.024	0.068+	0.091*	0.118**	0.050
Control	0.244***	0.261***	0.338***	0.202***	0.094**	0.171***	0.135***	0.145***
Challenge	0.052	0.097**	0.103**	0.061+	0.058	−0.011	0.015	0.056
Constant	0.221**	0.189**	0.102	0.148	0.118	0.162+	0.125	0.132
R-Square	0.285	0.277	0.312	0.279	0.202	0.230	0.240	0.238
Adj. R-Square	0.194	0.184	0.224	0.186	0.100	0.132	0.142	0.140
Classrooms	1892	1892	1892	1892	1892	1892	1892	1892
Teachers	1275	1275	1275	1275	1275	1275	1275	1275
Schools	208	208	208	208	208	208	208	208

Note: Each school has a separate intercept term and all regressions include level (i.e., elementary versus middle school) and subject (i.e., ELA versus math) indicator variables.

Standard errors adjusted for clustering by teacher.

Two-tailed significance indicators: + 0.10; * 0.05; ** 0.01; *** 0.001.

TABLE 4.A.4. **Regressions Behind Figure 4.3**

	Value Added	Happy in Class	Effort in Class	Inspired Re: College	Value Added	Happy in Class	Effort in Class	Inspired Re: College
Column	**1.**	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**
FfT Instruction	0.063+	0.137***	0.095**	0.118***				
FfT Class Environment	0.127***	0.069+	0.110***	0.057+				
7Cs Support					−0.120***	0.802***	0.335***	0.600***
7Cs Control					0.244***	0.121***	0.125***	−0.004
7Cs Challenge					0.148***	−0.115***	0.238***	0.135***
Subject and Level Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Intercepts	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.005	0.137	0.061	0.041	0.057	−0.01	−0.01	−0.034
R-Square	0.214	0.163	0.169	0.189	0.203	0.639	0.435	0.532

Adj. R-Square	0.114	0.055	0.062	0.084	0.136	0.608	0.378	0.492
Classes	1938	1868	1868	1868	3002	2961	2961	2961
Teachers	1307	1261	1261	1261	1920	1902	1902	1902
Schools	212	207	207	207	226	225	225	225

Note: Each school has a separate intercept term and all regressions include level (i.e., elementary versus middle school) and subject (i.e., ELA versus math) indicator variables.

Standard errors adjusted for clustering by teacher.

Two-tailed significance indicators: + 0.10; * 0.05; ** 0.01; *** 0.001.

TABLE 4.A.5. **Regressions Behind Figure 4.5**

	Control	Challenge	Managing student behavior	Establishing culture for learning	Managing classroom procedures	Creating an environment of respect and rapport	Engaging students in learning
Care	–0.110**	–0.079*					
Clarify	0.340***	0.403***					
Captivate	0.203***	–0.018					
Confer	0.227***	0.184***					
Consolidate	0.051	0.374***					
Communicating with Students			0.291***	0.300***	0.319***	0.335***	0.212***
Using Assessment in Class			0.173***	0.264***	0.167***	0.189***	0.307***
Using Questioning and Discussion Techniques			0.068*	0.205***	0.099**	0.161***	0.286***

Subject and Level Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Intercepts	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.144*	0.114*	0.032	0.053	0.102	0.135*	0.037
R-Square	0.503	0.708	0.389	0.558	0.424	0.490	0.564
Adj. R-Square	0.461	0.683	0.311	0.502	0.351	0.425	0.508
Classes	3004	3004	1939	1939	1939	1939	1939
Teachers	1921	1921	1308	1308	1308	1308	1308
Schools	226	226	212	212	212	212	212