ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

ICPSR 34771

Measures of Effective Teaching: 1 - Study Information

Bill and Melinda Gates Foundation

User Guide

Inter-university Consortium for
Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

Terms of Use

Information about Copyrighted Content

ICPSR

USER GUIDE TO MEASURES OF EFFECTIVE TEACHING LONGITUDINAL DATABASE (MET LDB)

Produced for Inter-University Consortium for Political and Social Research
The University of Michigan
August 2018

Mark White
Brian Rowan

Revised and expanded by
George Alter
Lindsay Blankenship
Christopher Greene
Stephanie Windisch

Table of Contents

1.0 OVERVIEW 5
1.1 The MET Project 5
1.2 Using this Guide 6
1.3 Revisions and Additions in 2018 6
2.0 EDUCATION RESEARCH AND POLICY ISSUES ADDRESSED BY THE MET STUDY 7
3.0 RESEARCH QUESTIONS ADDRESSED BY THE MET STUDY 8
4.0 DESIGN OF THE CORE MET STUDY 11
4.1 Core Study Design: Year One 11
4.1.1 Sampling 11
4.1.2. Focal Grades and Subjects. 11
4.1.3 Classroom Video and Video-Scoring 14
4.1.4 Other Year One Data Collections 15
4.2. Year Two Study Design 15
4.2.1 Year Two Randomization. 16
4.2.2 Full Year Two Teacher Sample 16
4.2.3 Year Two Teacher Randomization Sample 16
4.2.4 Year Two Classroom Video and Video-Scoring 17
4.2.5 Student Test Score Data 17
4.2.6. Student Survey Data 17
4.2.7 Dropped Instruments and New Instruments 17
4.3 MET Sub-Studies 18
4.3.1 Value-Added Measures Using District Administrative Data 18
4.3.2 Phase 1 and Phase 2 Scoring of Teacher Video Data 18
4.3.3 Survey of the Enacted Curriculum (SEC) 18
4.3.4 UTeach Observation Protocol (UTOP) 19
4.3.5 National Board for Professional Teaching Standards (NBPTS) 19
5.0 ADDITIONAL INFORMATION ON MET SAMPLING PROCESSES AND REALIZED SAMPLES 20
5.1 Recruitment Processes 20
5.1.1 Recruitment of Participating Districts 20
5.1.2 Recruitment of Participating Schools 20
5.1.3 Recruitment of Participating Teachers 21
5.1.4. Recruitment of classrooms for the observation component of the study 21
5.1.5 Recruitment of Students 22
5.2 Realized Samples 23
5.2.1 Realized samples versus national populations 23
5.2.2. Realized teacher samples versus district populations 24
5.2.3 Non-compliance rates for randomization 24
6.0 ADDITIONAL INFORMATION ON DATA COLLECTION, INSTRUMENTS AND MEASURES 26
6.1.1 SAT-9 open-ended (OE) reading assessment 26
6.1.2 Balanced assessment of mathematics (BAM) 26
6.1.3 ACT QualityCore 27
6.1.4 Construction of value-added measures 28
6.2 The Student Perceptions (TRIPOD) Survey 29
6.3 Classroom Videos and Video Scoring Processes 29
6.3.1 Video data collection 29
6.3.2 Raters and rater training 30
6.3.3 Phases of scoring 30
6.3.4 Assignment of raters to videos. 30
6.3.5 Scoring procedures 31
6.3.6 Double-scored videos 31

6.4.1 The Classroom Assessment Scoring System (CLASS) 32
6.4.2 Framework for Teaching (FFT) 33
6.4.3 Mathematical Quality of Instruction (MQI lite) 34
6.4.4. Protocol for Language Arts Teaching Observation (PLATO Prime) 34
6.4.5 Quality of Science Teaching (QST) 35
6.5 Content Knowledge for Teaching Assessment (CKT) 36
6.6 Teacher Working Conditions Survey (Year One) 36
6.7 Teacher Survey (Year Two) 37
6.8 Principal Survey 37
7.0 STRUCTURE OF THE CORE FILES (ICPSR 34414) 38
7.1 Data Not Included in the MET Core Files 38
7.2 Overview of the Structure of the MET LDB Core Files 38
7.3.2 School organization and demographics 40
7.3.3 Data on/from school principals 40
7.3.4 Principal survey data 40
7.4 Teacher File 40
7.4.1 ID Variables 40
7.4.2 Teacher study status 41
7.4.3 Teacher personal/professional Background 41
7.4.4. Teacher working conditions survey 41
7.4.5 MET teacher survey 41
7.4.6 CKT assessment 41
7.4.7 Principal ratings of teacher effectiveness 42
7.5 Section File 42
7.5.1 ID Variables 42
7.5.2 Section study status 42
7.5.3 Student Composition 42
7.5.4 Value Added Scores 42
7.5.5 Classroom Observation Scores 43
7.5.6 Student Survey Aggregates 43
7.6 Student by Section File 43
7.6.1 ID Variables 44
7.6.2 Student Study Status 44
7.6.3 Student Demographic Data 44
7.6.4 Student Test Scores 44
7.6.5 Student Survey Data 45
7.6.6 Using the Student by Section File 45
7.7 Observation Scores Files (Segment Level Files) 45
7.7.1 ID Variables 45
7.7.2 Observation Scores 46
7.7.3 Score Descriptors 46
7.7.4 Audio/Video Quality Scores 46
7.8 Video Files 46
7.8.1 Video Information File 46
7.8.2 Availability of Videos. 46
7.8.3 MET Extension Project Videos. 47
7.8.4 Video Format 47
8.0 DATA COLLECTIONS IN THE MET LDB 48
9.0 DISTRICT-WIDE FILES AND TEACHER VALUE-ADDED 50
9.1 Overview of the District-Wide Files 50
9.2 Data files, missing values, test years, and grade levels 50
9.3 Computing value-added estimates 51

9.4 Linking students and teachers in the District-Wide Files 52
9.5 MET student indicators 52
10.0 COMPARISON OF PHASE 1 AND PHASE 2 VIDEO SCORING 53
11.0 OBSERVATION SCORE CALIBRATION AND VALIDATION FILE (ICPSR 37090) 60
APPENDIX A. ID TYPES AND FREQUENCIES 61
A.1 Merging Files in the MET LDB 61
APPENDIX B. DESCRIPTION OF RANDOMIZATION PROCESS 83
APPENDIX C. BIOLOGY LESSON SELECTION FOR TEACHERS 87
APPENDIX D. TEACHER WORKING CONDITION CONSTRUCTS 89

1.0 OVERVIEW

1.1 The MET Project

The *Measures of Effective Teaching* project was the largest study of classroom teaching ever conducted in the United States. Supported by a grant from Bill and Melinda Gates Foundation, MET researchers collected a variety of indicators of teaching quality over a two year period (AY 2009-2010 and AY 2010-2011) in the classrooms of more than 2500 fourth- through ninth-grade teachers working in 317 schools located in six large school districts in the United States.

The data collected on teachers and their teaching included: (a) measures of students' achievement in each teacher's classroom drawn from state-administered assessments and supplemental achievement tests; (b) surveys of students in each teacher's classes; (c) video-recorded lessons taught by teachers and scored by independent observers using multiple classroom observation protocols; (d) assessments of each teacher's pedagogical and content knowledge for teaching; and (e) two different teacher surveys. In addition, principals of the schools where teachers worked also completed a survey and other administrative data on schools, teachers, and students are available for analysis.

The MET Study addressed several related research questions:

- How reliable and valid are the specific measures of teaching effectiveness under study? Do the various measures identify distinctive dimensions of teaching effectiveness, and if so, what dimensions are identified? What measures of effective teaching are empirically related to student learning gains?

- What does effective teaching look like, and how does it compare to less effective teaching? For example, what is the distribution of teacher scores on measures of effective teaching, and how much difference is there in teacher knowledge scores, teaching practice scores, and student outcome scores among teachers at different points in the distribution of measures of effective teaching?

- Can multiple sources of data on teachers and their teaching be combined to develop a set of fair, valid, and reliable indicators of teaching quality for use in teacher evaluation systems intended to rank teachers for personnel decision making and to promote teachers' professional learning and development?

To address these questions, the MET Study studied the intact classes of participating teachers during the 2009-2010 school year then randomly assigned teachers to classes of students in the 2010-2011 school year. In this design, the 2009-2010 study of intact classrooms was devoted to constructing measures of teaching effectiveness, to assessing the psychometric properties of various measures of teaching effectiveness, and to using correlational methods to assess empirical relationships among measures. The second-year (or randomization) portion of the study collected much of the same data on teachers as in Year One and was designed to make causal inferences about relationships among indicators of teaching quality.

1.2 Using this Guide

This report describes: (1) the educational research and policy context for the study; (2) research questions addressed by MET researchers; (3) the core study design and sub-studies; (4) the realized study sample; (5) additional information on study instruments and derived measures; and (6) the data files available for secondary analysis.

Most of this Guide introduces the MET LDB Core Files, which include the data most consistent with the underlying research design of the MET Project. **ICPSR strongly recommends that researchers base their analyses on the Core Files whenever possible. Even researchers who will not use the Core Files should read the description of those files carefully to understand the many dimensions of the MET LDB collection.**

1.3 Revisions and Additions in 2018

With additional support from the Bill and Melinda Gates Foundation ICPSR has been able to update and add to the MET LDB. The main changes are:

- District-wide data used to compute "value-added" measures have been added for school years 2011-12, 2012-13, and 2013-14. Variable names have been harmonized across all years and districts. See section 9.2 Computing value-added estimates.

- All district-wide data files have been updated with "ICPSR_GLOBAL_IDs" that link students and teachers across all data files in all years. Some students and teachers can be followed across six school years, 2008-09 to 2013-14. See 9.3 Linking students and teachers in the District-Wide Files.

- Student and teacher IDs in the Randomization File have been updated using "ICPSR_GLOBAL_IDs." This means that students and teachers who were not in the MET Project can now be linked to the district-wide data. See ICPSR Study # 34771 - Study Information in section 8.0 Data Collections in the MET LDB.

- A new Observation Score Calibration and Validation data file has been added to the MET LDB. The raters who scored MET videos were constantly monitored and evaluated. The result is a database of more than 2.4 million scored items from five observation instruments: Framework for Teaching (FFT), Classroom Assessment Scoring System (CLASS), Mathematical Quality of Instruction (MQI), Protocol for Language Arts Teaching Observations (PLATO), and Quality of Science Teaching (QST). These data are provided to encourage psychometric research on rater error. See section 11.0 Observation Score Calibration and Validation File.

2.0 EDUCATION RESEARCH AND POLICY ISSUES ADDRESSED BY THE MET STUDY

Several decades of research has shown that classroom-to-classroom variation in students' academic achievement is substantial, even among students at the same school. In a typical study, the classrooms to which students are assigned account for somewhere between 7-21% of the variance in students' academic achievement gains over a one-year period, which in some studies is more than the percentage of variance accounted for by the school that students attend (e.g., Murnane, 1975, Sanders and Horn, 1998; Rowan, Correnti, & Miller, 2002; Nye, Konstantopoulos, & Hedges, 2004; Rockoff, 2004; Rivkin, Hanushek, & Kain, 2005; Gordon, Kane & Staiger 2006).

At the time of the MET Study, researchers were asking two questions about classroom-to-classroom variation in student achievement gains. Some researchers questioned whether such variation represented true "teacher effects" on student achievement or whether such variation was due to other factors, in particular, measured and unmeasured differences in student characteristics that vary across classrooms (e.g., Rothstein, 2010). However, other researchers assumed that classroom-to-classroom variation in student achievement gains was (in fact) due to teacher effects, and these researchers were inquiring into the specific characteristics of teachers and their teaching that might explain such variation. At the time of the MET Study, researchers were entertaining several scientifically credible explanations, including differences among teachers in professional knowledge for teaching, as well as differences in the nature and quality of classroom instruction, including patterns of classroom organization and management, the intellectual challenge of classroom work, and the emotional tenor of student-teacher relationships (e.g., Hill, Rowan, and Ball, 2005; Danielson, 2006; Pianta and Hamre, 2009; Ferguson, 2009).

Education policy makers also were interested in classroom-to-classroom differences in student achievement. At the time of the MET Study, policy makers were citing findings on teacher effects to justify an increased focus on teacher quality. An important federal initiative in this area was *No Child Left Behind* (NCLB), which required schools to place "highly qualified" teachers in every classroom. However, NCLB defined highly qualified teachers in terms of professional certification and degree attainment, even though research had not found these factors to be highly correlated to student achievement gains. The Race to the Top (RTTT) initiative (launched in 2009) changed the focus of teacher quality initiatives at both the federal and state levels. Instead of focusing on teacher qualifications, the RTTT initiative focused on "improving teacher…effectiveness based on performance." To achieve this aim, the RTTT initiative called for school systems to design and implement rigorous, transparent, and fair teacher evaluation systems that were meant not only for use in personnel evaluation and decision making, but also to identify teachers' professional learning and development needs.

3.0 RESEARCH QUESTIONS ADDRESSED BY THE MET STUDY

In this environment, the MET Study had several goals. One goal was to collect the kinds of data required to develop multiple measures of effective teaching for use in education research and practice. Among the effectiveness indicators MET researchers collected was data on the achievement gains of students in each teacher's classroom(s). Importantly, these gains were calculated for scores from state assessments and for scores on alternative assessments administered by MET researchers. Using these data, MET researchers developed several "value-added" measures of teaching effectiveness for participating teachers. In addition, MET researchers video recorded each participating teachers' classroom instruction on multiple occasions. These videos were then coded by trained raters using several different classroom observation protocols to yield yet another set of measures of teaching effectiveness. A third set of measures of teaching effectiveness collected by MET researchers sought to measure teachers' knowledge for teaching. Here, the MET Study developed an assessment of what has been called teachers' pedagogical content knowledge. A final set of measures used in the MET Study came from student surveys administered in participating teachers' classrooms. These surveys asked students to report on a variety of dimensions of instruction and classroom climate. All of these measures of teaching effectiveness are described at later points in this report, and the reader also can find descriptions of these measures in a MET research report entitled *Gathering Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains* (Gates Foundation, 2012)

Once multiple measures of teaching effectiveness were collected, MET researchers conducted some very basic descriptive analyses. Although the MET Study does not include a nationally representative sample of teachers, MET data can be used to examine both central tendencies and variation in measures of effective teaching in a large sample of teachers at grades 4-9 working in six large urban school districts in the United States. The descriptive analyses conducted by MET researchers involved examining both central tendencies in teachers' scores on the various measures of effective teaching and an examination of score distributions that allowed MET researchers to quantify how much difference in observation scores existed among teachers at different points in each observation measure's score distributions. The results of these analyses were reported in *Gathering Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains* (Gates Foundation, 2012: pp. 28-29).

Beyond these basic analyses, MET researchers also analyzed the psychometric properties of different measures of teaching effectiveness. For example, using first-year data from the study, MET researchers explored the reliability of value-added measures by looking at the correlations of the same value-added measures across the different class sections taught by middle school teachers in the same year and by correlating teachers' value-added scores across years of the study. MET researchers also explored the reliability of measures of teaching effectiveness derived from classroom observation protocols. Here, they conducted generalizability studies that partitioned variance in scores due to observation days, class sections, observers, and error, and used these analyses to assess measure reliabilities. The results of these analyses are reported in *Gathering Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains* (Gates Foundation, 2012: pp. 34-40).

Once the psychometric properties of measures were established, MET researchers estimated empirical relationships among various measures of teaching effectiveness in the tradition of multi-trait/multi-method research. Here, for example, first-year MET data (from intact classrooms) were used to examine the correlations among different value-added measures and correlations among different classroom observation measures. These correlational analyses (and related factor analytic

studies) allowed MET researchers to identify various dimensions of teaching being measured by the different instruments and to assess the extent to which different measures were assessing the same or different dimensions of teaching effectiveness. The results of these analyses were reported in *Gathering Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains* (Gates Foundation, 2012: pp. 28-33).

MET researchers also used first-year data in a set of validity analyses that examined how measures of teaching based on classroom observations and student surveys were associated with student learning gains. In these analyses, researchers used first-year data to examine correlations of value added measures to measures derived from classroom observations and student surveys. However MET researchers wanted to move beyond these kinds of correlational analyses in order to make causal inferences about the nature of teaching effectiveness. This was not possible using first-year MET data, however, because in that data set, all of the measures of teaching effectiveness were gathered from the same classrooms at the same point in time. Under these conditions, it would be quite possible for correlations among value-added measures of teaching effectiveness and (for example) classroom observation measures to be due, not to stable underlying traits of teachers, but rather to unmeasured characteristics of the students in a teachers' classroom. To combat this problem, MET researchers took several steps. In one effort, they correlated first-year classroom observation measures to student achievement gains in teachers' prior year classrooms ("post"-dicting). In a second analysis, they took advantage of the fact that in MET first-year data, many participating middle school teachers taught more than one class section of students. MET researchers capitalized on this situation to separate teacher effects from student composition by examining empirical relationships among measures of teaching effectiveness across the different class sections taught by the same teacher in the same year. The results of these analyses are reported in *Gathering Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains* (Gates Foundation, 2012: pp. 43-50).

However, another way to sharpen causal inferences about the relationships of indicators of teaching effectiveness is to examine empirical relationships among measures across years. This form of analysis can be done with MET second-year data. Importantly, prior to Year Two of the MET Study, teachers were randomly assigned to student groups to prevent selection bias from operating in the sorting of students to teachers—as would happen, for example, if better-performing students systematically found their way to some teachers, while worse-performing students systematically found their way to other teachers. With selection bias, the same un- or mis-measured characteristics of students could account for year-to-year correlations among measures of teaching effectiveness. However, with randomization, this explanation can be ruled out. As a result, the causal inference that an "effective" teacher in Year One produced higher learning gains among her assigned students in Year Two becomes plausible. The approach taken by MET researchers to issues of causal inference in research on effective teaching will be described in a forthcoming MET report (to be posted on the MET web site: www.metproject.org).

A final goal of the MET Study was to connect basic research on teaching effectiveness to education policy and practice, particularly policy and practice in the area of teacher evaluation. As discussed earlier, an important emphasis in the federal government's Race to the Top initiative has been to promote the development of teacher evaluation systems that identify effective teachers on the basis of their impact on student achievement and that provide feedback to help teachers improve. A problem is that value-added measures of teaching effectiveness, which are based on test score outcomes, allow teachers and administrators to compare the impact of specific teachers on student learning, but they do not provide information about what teachers might do to improve students' learning outcomes. That is why the development of multiple indicators of teaching effectiveness is

important. The MET Study's approach to combining multiple measures of teaching effectiveness and assessing reliability and validity using first-year MET data is discussed in *Gathering Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains* (Gates Foundation, 2012: pp. 45-55). The approach taken by MET researchers with second year data will be described in a forthcoming MET report (to be posted on the MET web site: www.metproject.org).

4.0 DESIGN OF THE CORE MET STUDY

This section provides a brief overview of what this *User Guide* calls the "core" elements of the MET Study design (the MET "core study"). As the reader will see, the MET Study conducted a multifaceted, longitudinal study that collected many different forms of data on large numbers of teachers using a complex research design. For this reason, **researchers are encouraged to focus on those aspects of the study that are consistent across time and that have been central to the work of MET researchers.**

Section 4.0 of the *User Guide* presents a very brief overview of these core study elements. More detail on various aspects of the study design, instrumentation, and data collection procedures are then provided in subsequent sections (5-7) of this *User Guide*.

4.1 Core Study Design: Year One

We provide here a brief overview of the sampling and data collection procedures used in Year One (AY 2009-2010) of the MET Study. Tables 1 and 2 offer an "at-a-glance" listing of these study elements. The remainder of this section provides a bit more detail.

4.1.1 Sampling. The MET Study began with a process of "opportunity" sampling that took place over the period July - November 2009 and that resulted in six, large school districts volunteering to participate in the study. The process of opportunistic sampling then continued as elementary, middle, and high schools within each district were recruited into the study. Once schools were recruited, opportunity sampling continued as teachers (at targeted grade levels and subject areas) within these schools volunteered for the study. A detailed discussion of MET Year One sampling (as well as the resulting samples of districts, teachers, and students that resulted from this process) is presented in Section 5.0 of this *User Guide*. As discussed there, the sampling process resulted in 2,741 teachers from 317 schools in six large school districts being recruited into the first year of the study (AY 2009-2010).

4.1.2. Focal Grades and Subjects. The MET Study was designed to study issues of teaching effectiveness within a focal set of grades and subjects:

Elementary Grades (4 & 5). At the elementary grades, the MET Study focused on the teaching of English Language Arts (ELA) and Mathematics at grades 4 and 5. Of the 4th and 5th grade teachers recruited into the study, the vast majority were subject-matter generalists who taught English Language Arts (ELA) and Mathematics to a *single* class of students. However, a smaller number of teachers at grades 4 and 5 were either subject matter specialists (who taught ELA or Mathematics to more than one class section of students) or teachers who only volunteered to have their teaching of a single subject studied.

TABLE 1: Samples of the MET Study by Year			
Sampling Plan	**Full Sample All Year One Teachers (AY 2009-2010)**	**Core Study Sample = All Teachers Present in Year Two (AY 2010-2011)**	**Randomization Sample = Teachers Randomized in Year Two (AY 2010-2011)***
Districts	**6 districts participate**	**6 districts continue**	**6 districts continue**
Schools	**Opportunity sampling (grade by subject exchange groups required) 317 schools participate**	**310 schools continue in the study.**	**284 schools with teachers randomly assigned to classes continue in study.**
Teachers	**Opportunity sampling (teacher must be in exchange group at school). 2,741 teachers participate**	**2,086 teachers continue in the study.**	**1,559 teachers randomly assigned to classes during summer continue in study.**
Class sections	**Opportunity sampling (specialist teachers nominate class sections for study). 4,497 class sections in study**	**1,909 class sections present in second year of the study**.**	**1,379 class sections (one per teacher) randomly assigned by MET researchers.**

*** It should be noted that the randomization sample is a sub-sample within the core study sample. Both of these samples are, of course, sub-samples of the full sample.**

**** A teacher who was randomly assigned during the summer, but dropped out of the study between random assignment and the start of the school year (184 teachers), will not have a section present in Year Two. Otherwise, teachers in the core study sample and the randomization sample will almost always have one Year Two section.**

TABLE 2: Core Elements of the MET Study by Year*		
District Administrative Data	**Year One (AY 2009-2010)****	**Year Two Sample (AY 2010-2011)****
School	**Data on grade configuration, enrollment size, student composition**	**No new data**
Teachers	**Data on personal and professional background**	**No new data**
Students	**Data on prior state test scores, current test scores (grades 4-8), sex, ethnicity, free lunch participation, program participation**	**Data on prior state test scores, current state test scores, sex, ethnicity, free lunch participation, program participation**
Classroom Videos	**Year One (AY 2009-2010)****	**Year Two Sample (AY 2010-2011)****
Subject Matter Generalists	**Each teacher video recorded on four days. Each day includes two video sessions: one ELA, one Mathematics**	**Each teacher video recorded on four days. Each day includes two video sessions: one ELA, one Mathematics**
Subject Matter Specialists	**Each teacher video recorded on two days. Each day includes video session in two sections**	**Each teacher video recorded on four days. Each day includes video session in one section**
Classroom Video Scoring	**Year One (AY 2009-2010)****	**Year Two Sample (AY 2010-2011)****
Subject Matter Generalists	**Each video coded with CLASS and FFT. ELA sessions also coded with PLATO. Math sessions coded with MQI.**	**Each video coded CLASS and FFT. ELA sessions also coded with PLATO. Math sessions coded with MQI.**
Subject Matter Specialists	**English and Math Sessions are coded with CLASS and FFT. English sessions also coded with PLATO. Math sessions coded with MQI. Biology sections coded with QST**	**English and Math Sessions are coded with CLASS and FFT. English sessions also coded with PLATO. Math sessions coded with MQI. Biology sections coded with QST**
MET Student Data Collections	**Year One (AY 2009-2010)****	**Year Two Sample (AY 2010-2011)****
Grades 4-5	**SAT-9 open-ended reading BAM**	**SAT-9 open-ended reading BAM**
Grades 6-8	**SAT-9 open-ended reading BAM**	**SAT-9 open-ended reading BAM**
Grade 9	**ACT Quality Core English Grade 9 ACT Quality Core Algebra I ACT Quality Core Biology**	**ACT Quality Core English Grade 9 ACT Quality Core Algebra I ACT Quality Core Biology**
School Personnel Surveys	**Year One (AY 2009-2010)****	**Year Two Sample (AY 2010-2011)****
Teachers	**Teacher Working Conditions Survey**	**MET Teacher Survey CKT Assessment**
Principals		**MET Principal Survey**

*** Abbreviated terms described in text.**

**** Year one measures were given to the full sample. Year two measures, with the exception of video scoring, were all given to the core study sample and the randomization sample. Video scoring focused only on the randomization sample.**

Middle Grades (6-8). At the middle grades, the study focused on the teaching of ELA and Mathematics in grades 6-8. Of the middle school teachers who volunteered for the study, about half were teachers of ELA in grades 6-8, and the other half were teachers of Mathematics at these grades. In addition, a handful of sixth grade teachers were subject-matter generalists who taught ELA and Mathematics to a single class of students.

High School (Grade 9). At the high school level, the study focused on the teaching of 9th grade English, 9th grade Algebra I, and 9th grade Biology. Of the 9th grade teachers who volunteered for the study, about a third were teachers of 9th grade English, another third were teachers of 9th grade Algebra, and another third were teachers of 9th grade Biology.

4.1.3 Classroom Video and Video-Scoring. One of the most important features of the MET Study was the video recording of a sample of each participating teacher's classroom instruction and the subsequent scoring of these video recorded sessions using multiple observation protocols. The process of video recording and scoring videos is discussed in greater detail in Section 6.3 of this *User Guide*. Here, we present only a very brief overview of the process. The video recording process differed for generalist teachers, who taught most subjects to a single group of students, and specialist teachers, who taught a single subject to multiple groups of students.

Subject Matter Generalists. As discussed, most elementary teachers in the MET Study (and a handful of 6th grade teachers) were subject matter generalists—that is, they taught multiple subjects to a single group of students. During Year One of the study, these teachers were video recorded on four separate days, and on each day, MET researchers video recorded these teachers as they taught both ELA *and* Mathematics[1]. Once videos were recorded, the videos were scored by trained raters using procedures described at a later point in the *User Guide*. The important point, for now, is that each elementary school generalist teacher produced 4 ELA videos and 4 Mathematics videos for scoring. Each of these videos was scored using two general teacher observation protocols: the upper elementary version of the Classroom Assessment and Scoring System (CLASS), and the Framework for Teaching (FFT) protocol. In addition, each ELA video session was scored using the subject-specific Protocol for Language Arts Teaching Observation (PLATO), and each mathematics video session was scored using the subject-specific observation protocol known as Mathematical Quality of Instruction (MQI). Moreover, about 5% of these videos were double-scored (i.e., scored by two raters).

Subject Matter Specialists. Most teachers in 6th-9th grade and a handful of 4th and 5th grade teachers were specialist teachers. These teachers taught the same subject to multiple groups of students. In Year One of the study, these specialist teachers were video recorded as they provided instruction in *two* of their class sections. Each class section was video recorded on two days during the school year, with both class sections being taped on the same day, giving four total videos for each teacher. Then, each math and English video session was scored using both the general protocols (CLASS-upper elementary for grades 4-6 or CLASS-secondary for grades 7-9, and FFT) and the relevant subject-specific protocol (PLATO for ELA sections, MQI for mathematics sections). Biology video sessions were only scored using the QST instrument. About 5% of these videos were double-scored (i.e., scored by two raters)[2].

[1] Some generalist teachers only volunteered a single subject for the study. The videotaping procedure they followed differed only in that only the one subject was video recorded.
[2] For QST, the double scoring rate was 10%.

4.1.4 Other Year One Data Collections. During Year One of the MET Study, a large amount of data beyond classroom observations also was collected by MET researchers. These data collections were designed to provide additional measures of teaching effectiveness as well as greater detail on the schools, teachers, and students in the study. These data collections are discussed in greater detail at later points in this *User Guide*. This section simply provides a brief overview of these data.

District Administrative Data. Each district provided administrative data on the schools, teachers, and students in the study. Administrative data on schools included measures of a school's enrollment size for grades in the MET Study, grade configuration, and student composition. Administrative data on teachers included measures of a teacher's sex, ethnicity, years of teaching experience, and degree status. Administrative data on students included measures of students' sex, ethnicity, free lunch status, program participation status, and multiple years of scores on state achievement tests. Data on schools, teachers, and students were linked so that it is possible using MET Study data to identify which students were in a particular teacher's classes at multiple time points during the MET Study.

Student Test Score Data. MET researchers also used a variety of student assessment instruments to measure student learning and develop "value-added" measures of teacher effects on student learning. At grades 4-8, student learning was measured by state assessments (typically in reading and mathematics) as well as by two assessments administered directly by MET researchers. One of these "supplemental" assessments was the SAT-9 Open-Ended Reading Assessment, another was the Balanced Assessment in Mathematics (BAM). At Grade 9, state assessments were not available to measure student learning in most districts. Instead, MET researchers administered ACT Quality Core "end-of-course" assessments for Algebra I, English 9, and Biology. At a later point in this *User Guide*, we describe these assessments in more detail and discuss how they were used to estimate value-added scores of teaching effectiveness for teachers in the study.

Student Survey Data. In Year One of the study, MET researchers also administered a version of the Student Perception (or Tripod) Survey to students in the focal class sections under study. This survey asked students to rate seven dimensions of classroom instruction as they experienced it in the particular MET class section they attended. Two versions of the survey were administered, one for students in grades 4-5, and one for students in grades 6-9. The versions differed mostly in the wording of questions.

Teacher Working Conditions Survey. In Year One of the study, MET researchers also administered a Teacher Working Conditions Survey to all participating MET teachers and to all other teachers in MET schools. This survey had more than 200 items asking teachers to report on many different features of their school, using many items that were borrowed from previous school surveys conducted in the United States. Survey items were intended to measure aspects of school policies and procedures, supports for technology, professional development and learning, school improvement processes and planning, teacher participation in decision making, school personnel practices, and teachers' beliefs about various aspects of teaching and learning.

4.2. Year Two Study Design

As noted earlier, the MET Study was longitudinal and aimed to study teachers over a two-year period. As discussed earlier, Year Two was designed to feature a randomization component in which classes of students were randomly assigned to teachers. As discussed in earlier portions of this *User Guide*, this feature of the study was implemented in order to improve causal inferences about teacher

effectiveness. Beyond this randomization component, Year Two of the MET Study included data collections that differed somewhat from Year One data collections. All teachers participating in Year Two of the study were observed (using similar video recording and scoring procedures used in Year One), and some data collections were continued (including collection of district administrative data on students, administration of student assessments, and administration of the student survey). However, an assessment of teachers' knowledge for teaching was administered in Year Two of the study, a different teacher survey was administered in Year Two, and a new survey was administered to principals of MET schools in Year Two. These Year Two procedures are now described.

4.2.1 Year Two Randomization. Section 5.1.3 of the *User Guide* discusses in detail how teachers were selected for participation in Year Two of the MET Study. Therefore, in this section, we provide only a very brief overview of this process. The process began in Year One of the study when MET researchers initially recruited schools into the study. As discussed in Section 5.1.3 of the *User Guide*, only schools that had more than one teacher in a grade teaching the same subject (a grade/subject combination) were allowed to enter the study. The reason for this was that MET researchers intended to randomly assign classes of students to teachers in Year Two of the study, and to do so, they needed to form what they called "exchange" groups (or "randomization blocks") of teachers within the same school to whom classroom groups could be assigned. At least two members of an exchange group had to be teaching at the same school at the time of randomization for teachers to be randomized and included in the core study.[3]

4.2.2 Full Year Two Teacher Sample. There were additional sources of attrition in the teacher sample in Year Two of the study. Some teachers were lost when their school dropped out of the study (11 schools; 60 teachers). Additionally, individual teachers dropped out when they left their school or district, began teaching a different subject or grade, lost interest in the study, or became ill. Overall, the Year Two sample of teachers included 2086 teachers in 310 schools. Of the 582 4th and 5th grade teachers in Year Two, the majority continued to be subject-matter generalists who taught English Language Arts (ELA) and Mathematics to a single class of students, although the sample also included a small number of subject matter specialists (who taught ELA or Mathematics to more than one class section of students) and teachers who volunteered only to have their teaching of a single subject be studied. Of the 841 middle grades teachers in Year Two, about half continued to be teachers of ELA in grades 6-8, and the other half teachers of Mathematics at these grades. Of the 479 9th grade teachers in Year Two, about a third were teachers of 9th grade English, another third were teachers of 9th grade Algebra I, and another third were teachers of 9th grade Biology.

4.2.3 Year Two Teacher Randomization Sample. The goal of the second year of the study was to randomly assign teachers to classrooms. However, not all teachers could be randomly assigned. The randomization sample is a sub-sample of all teachers present in Year Two. It consists of 1559 teachers in 284 schools. The 527 teachers in the Year Two sample, but not in the randomization sample, could not be randomized because all the other teachers in their exchange group had left the study or their school decided that it would no longer consent to randomization.

[3] In NYC, 14 new teachers were recruited so that existing teachers could be randomized and remain in the study.

4.2.4 Year Two Classroom Video and Video-Scoring. MET researchers used slightly different procedures for video recording and scoring classroom sessions across years of the study[4]. The procedures for subject matter generalists remained the same in Year Two of the study. That is, subject matter generalists were video recorded on four days, with each day producing both an ELA and a Mathematics video, and these videos were once again scored using CLASS and FFT (the general observation protocols) and the relevant subject-specific protocol (PLATO or MQI). However, video recording was different for subject-matter specialists in Year Two. Specialists only had one section included as part of the study and that single section was observed on four different days. Scoring followed the same process as Year One, with videos subsequently scored using CLASS, FFT, and the relevant subject-specific protocol (PLATO or MQI). Biology videos were again only scored with the QST.

4.2.5 Student Test Score Data. Procedures for student testing remained the same across years of the MET Study. Once again, in Year Two, researchers measured student learning in grades 4-8 using state assessments (in reading and mathematics) and by two assessments administered directly by MET researchers—the SAT-9 Open-Ended Reading Assessment, and the Balanced Assessment in Mathematics (BAM). At Grade 9, MET researchers once again administered ACT Quality Core assessments for Algebra I, English 9, and Biology.

4.2.6. Student Survey Data. In Year Two, MET researchers also administered the Student Perception (or Tripod) Survey to students in the focal class sections under study. Scales were added to this survey to measure characteristics of the classroom beyond the 7Cs. These are described in detail in later sections.

4.2.7 Dropped Instruments and New Instruments. In Year Two of the study, district administrative data were collected on students new to the study. However, new administrative data on teachers and schools were not collected. In addition, the Teacher Working Conditions Survey was not re-administered. Instead, three new surveys were administered in Year Two of the study:

4.2.7.1 The Content Knowledge for Teaching Assessment (CKT). In Year Two of the study, MET researchers administered a web-based assessment of teachers' pedagogical content knowledge for teaching. Multiple forms of this assessment were created, and scores on the different forms cannot be equated. Separate assessments were created and administered for grades 4-6 ELA, grades 7-9 ELA, grades 4-5 Mathematics, grades 6-8 Mathematics, and Algebra I. In the MET Study design, subject matter generalists teaching ELA and Math in 4th and 5th grade needed to take two assessments (one for ELA, one for Math). The Mathematics assessment for grades 4-5 was therefore administered in Fall, 2010 and the English Language Arts assessment was administered in Winter, 2011. Teachers in all other grades were administered only one assessment (in the subject they taught), and these assessments were administered in Winter of 2011[5].

4.2.7.2 The MET Teacher Survey. At the end of the Year Two (AY 2010-2011), teachers still participating in the MET Study also were administered a web-based MET Teacher Survey that asked them to report about various aspects of their work with principals and the teacher evaluation system in place at their school.

---

[4] The reader should note that MET researchers used a different procedure to score samples of videos in Year One of the study. However, all videos of teachers in the *core* randomization sample were coded using the same method across years.

[5] Four 6th grade generalist teachers took the assessment in both subjects during Winter 2011.

4.2.7.3 The MET Principal Survey. A final, new instrument administered in Year Two was the MET Principal Survey. This survey was administered to all principals in MET schools at the end of the AY 2010-2011 school year. The survey contained questions about the school's current teacher evaluation policies, the training received by a principal on issues of teacher evaluation, the principal's comfort with teacher evaluation, the principal's perceptions of the effectiveness of the teacher evaluation system currently in place, and the principal ratings of the teaching effectiveness of up to 12 of the MET teachers in his or her school.

4.3 MET Sub-Studies

The data collections just described constitute what are here called the "core" study. The data collected in this core study will be consistent across years and can be used to explore all of the central questions relevant to the MET Study. But the data collections constituting the core study are *not* the only ones developed by MET researchers. Additional data collections, called "sub-studies" in this *User Guide*, are discussed here. The reader should note that data from these sub-studies are not included in the Core Files (ICPSR Study 34414), but they are available from ICPSR in other studies within the MET series.

4.3.1 Value-Added Measures Using District Administrative Data. The Core Files contain value-added measures (VAMs) of teaching effectiveness derived from district administrative files (using state achievement tests and statistical procedures developed by MET researchers and described at a later point in the *User Guide*). The administrative data used to compute VAMs are available in ICPSR Study 34798 "District-Wide Files, 2009-2014."

4.3.2 Phase 1 and Phase 2 Scoring of Teacher Video Data. MET Study researchers used different scoring procedures at different points in the study. Over the course of the study, video scoring was done in two phases. Scoring began with an initial focus on 2,000 videos from a subset of 4th-8th grade Year One teachers with complete data. These 2,000 videos (413 teachers; 739 classrooms) are described here as the "Phase 1"[6] sample. Teachers with complete data were stratified by grade, subject, and district then randomly selected to be a part of Phase 1. All Year One videos from selected teachers were included in the sample. Phase 2 observation scoring was applied to both Year One and Year Two videos of teachers in the randomization sample, who are included in the Core Files (14,580 videos). Phase 1 scoring procedures differed in several ways from the scoring in Phase 2. During Phase 1 scoring, raters used a different coding interface, coded different numbers of video segments, and were assigned to perform rating tasks differently from procedures used in the Phase 2 scoring. Approximately, 1,300 Year One videos were scored in both Phase 1 and Phase 2. The Core Files (ICPSR Study 34414) contain only Phase 2 scores. Scores from both Phase 1 and Phase 2 are available in ICPSR Study 34346 "Item-Level Observational Scores and Supplemental Test Files, 2009-2011." Summary scores from Phase 1 and Phase 2 are identified in the section-level analytical files (ICPSR Study 34309).

4.3.3 Survey of the Enacted Curriculum (SEC). The Survey of the Enacted Curriculum has been used in past studies to examine the content taught by teachers over the span of a school year and to measure its alignment to state curriculum standards. MET researchers administered this survey to a total of

[6] Within Phase 1, an initial set of videos was scored on the CLASS protocol during a "summer pilot." This was followed by "Plan B," which applied CLASS, MQI, PLATO, and FFT to 2,000 selected videos. These videos are sometimes called the "Plan B" videos in MET Project documents.

500 MET teachers at the end of Year One (AY 2009-2010). A random sample of 100 teachers from each of the following subject/grade combinations was surveyed: 4th grade ELA and Math, 8th grade ELA and Math, and 9th grade Biology. The SEC will be available from ICPSR in Fall 2013.

4.3.4 UTeach Observation Protocol (UTOP). The MET Study used a new mathematics and science observation scoring protocol to score the math videos in the Phase 1/Plan B sample. The new protocol was specifically developed to accurately capture classroom quality while valuing instructional strategies from inquiry-based learning to direct instruction. Training of raters and scoring of videos for UTOP was managed by the National Math and Science Initiative. No data from the UTOP observation protocol is included in the Core Files. Summary scores UTOP are available in the section-level analytical files (ICPSR Study 34309).

4.3.5 National Board for Professional Teaching Standards (NBPTS). The National Board for Professional Teaching Standards scored 700 one-hour MET videos using the National Board's Adolescent and Young Adulthood certificate rubrics for English Language Arts and Mathematics teaching. These data are not available in the Core Files. NBPTS is available in the section-level analytical files (ICPSR Study 34309).

5.0 ADDITIONAL INFORMATION ON MET SAMPLING PROCESSES AND REALIZED SAMPLES

Having provided a brief overview of the MET Study, we now provide additional detail on various study components. We begin in this section with a description of the MET sampling processes. Later sections provide more detail on additional study components.

5.1 Recruitment Processes

5.1.1 Recruitment of Participating Districts. MET researchers recruited districts into the study during the period July - November 2009. The recruitment targets were mostly large, urban districts that were receiving support from the Gates Foundation to develop human resource systems, although the recruiting effort was subsequently expanded to include other districts with which the Foundation had worked previously. The final selection of districts was based on a district's interest in the study, staff size sufficient to assure adequate numbers of participating teachers, central office support for the MET program, willingness and capacity to participate in all parts of the data collection process, and broader local political and union support for the project. At the end of recruitment, the following districts were selected for and participated in the study: Charlotte-Mecklenburg (NC) Schools, Dallas (TX) Independent School District, Denver (CO) Public Schools, Hillsborough County (FL) Public Schools, Memphis (TN) City Schools, and the New York City (NY) Department of Education. Pittsburgh Public Schools served as the project's pilot district, but no data from this district are included in the MET LDB. Each participating district received grant funding from the Gates Foundation that allowed for the hiring of at least one full-time district level project coordinator. Districts also participated in regular MET meetings with other districts and researchers.

5.1.2 Recruitment of Participating Schools. Within each recruited district, certain schools were excluded from participation in the study. These included special education schools, alternative schools, community schools, autonomous dropout and pregnancy programs, returning education schools, and vocational schools that did not teach academic courses. Also excluded from the study were schools that had team teaching or other structural features that made it impossible to assign responsibility for a student's learning to a single, specific teacher.

Once these exclusions were made, all schools in a district that contained the MET Study's target grades (4-9) were invited to participate. There were two exceptions. In Denver, no middle schools (grades 6-8) signed up for the program. Second, because the sample was complete except for middle school teachers by the time Dallas was recruited, only teachers in schools with grades 6-8 were recruited in Dallas. School recruitment within a district began immediately after that district agreed to participate in the study and concluded in February 2010. District coordinators led the school recruitment efforts in each district. The pool of schools for recruiting purposes was limited to schools that included any of grades 4, 5, 6, 7, 8, or 9, and that had at least three teachers assigned to one of the MET Study's focal subject/grade combinations. The focal grade/subject combinations were: grades 4-8 English/Language Arts (ELA), grades 4-8 Mathematics, Grade 9 English, Grade 9 Algebra I, and Grade 9 Biology. The requirement of three teachers teaching the same grade/subject combination effectively excluded small schools and most charter schools from the study.

Each eligible school was invited to participate in the study via a standard letter describing the project. The district coordinator then held informational meetings and took other steps to encourage principals to take part in the study. Schools were offered a variety of incentives to participate. Schools received $1,500 for use at the principal's discretion. Schools also received $500 a year to pay for a School Project Coordinator (NYC used grant savings to pay their coordinators $2,000 per year).

Districts also had some funds to provide minor incentives to schools such as money for school supplies. Last, the video recording equipment required for the classroom observation component of the MET Study was donated to the school at the conclusion of the study.

Principals who agreed to participate in the project completed an online sign-up form through which they provided general information about their school and the teachers in the school. They also committed to taking specific actions as part of the study. First, they agreed to participate in all aspects of data collection. Second, they agreed to randomly assign teachers to classrooms during Year Two of the study. This involved principals creating equivalent groups of students that could be randomly assigned to participating teachers.

5.1.3 Recruitment of Participating Teachers. Once a school principal agreed to participate in the study, all teachers assigned to teach MET Study focal grade/subject combinations were invited to participate in the study unless: (a) they were team teaching or looping, making it impossible to assign responsibility for the learning of a given student in a specific subject to that teacher[7]; (b) the teacher indicated that he or she was not planning to stay in the same school and teach the same subject the following year; or (c) there were less than two other teachers with the same grade/subject teaching assignments. This last restriction was put into place to assure that each teacher could be put into an "exchange" group for random assignment of classes to teachers in Year two of the study.

Once eligible teachers were identified, they were mailed a standard invitation to participate in the MET Study, and school principals, school-level coordinators, and the district coordinator actively encouraged them to participate. As volunteers at a school accumulated, they were placed into the exchange groups discussed above. Each "exchange group" consisted of a group of at least three teachers at a school who taught the same grade/subject combination to an equivalent group of students. Only teachers who could be placed in an exchange group were selected for the study, and schools that could not form at least 2 such exchange groups were eliminated from the study. In participating schools, teachers who were selected for study received a $1500 incentive for participating ($1000 at the beginning and $500 at the end of the study). Additionally, the districts were awarded small budgets to provide thank you gifts for teachers that participated.

5.1.4. Recruitment of classrooms for the observation component of the study. If a teacher agreed to participate in the study, the teacher also agreed to have his or her classroom instruction observed on several occasions during each school year of the study. This section of the overview describes how the classes to be observed were chosen. The procedure differed depending on whether a teacher was a subject matter generalist or subject matter specialist, and it differed across years. These procedures are now described.

5.1.4.1 Subject matter generalists. Most teachers in grades 4-5 and some teachers in grade 6 were subject matter generalists, that is, teachers who taught all academic subjects to a *single* class of students over the entire year. In Year One of the study, the class they taught was formed through normal processes of class scheduling at a given school, and the naturally-occurring class headed by each generalist teacher was designated as the classroom where that teacher's instruction was observed. In Year Two of the study, MET researchers developed a different sampling procedure. In that year, generalist teachers were still teaching all academic subjects to a single class of students.

[7] Classrooms with special education teachers were included in the sample if it was deemed that a single teacher could be assigned responsibility for student learning. However, it is impossible to know when and if this occurred.

But in Year Two of the study, MET researchers worked to randomly assign student groups to generalist teachers using the following procedure. Prior to the start of the school year, principals at participating schools formed grade level classes, attempting to make each class at a grade level as alike as possible in terms of student composition. Once this happened, MET researchers randomly assigned a pre-formed class to each teacher in a grade-level exchange group. Recall that exchange groups consisted of teachers at the same grade level (teaching the same subject), and for random assignment to take place within a school, there had to be at least two teachers in a grade-level exchange group. Teachers who could not be placed into an exchange group continued in the study, but were not in the randomization sample. These teachers participated with naturally formed sections.

5.1.4.2 Subject matter specialists. In grades 6-9, most teachers were subject matter specialists (i.e., they taught the same subject to more than one class section of students per day). Some teachers in grades 4-5 also were subject matter specialists. The selection of specialist teacher classes for observation was more complex than for subject matter generalists (who taught only one class group). Specialist teachers taught multiple class sections of students, but MET researchers decided to observe instruction in only <u>two</u> of these class sections. In Year One of the study, specialist teachers (no matter what the subject or grade level at which they taught) self-identified the two (naturally formed) class sections where instruction would be observed. In Year Two of the study, however, school principals were asked to identify multiple class periods when teachers from the same "exchange group" were teaching the same subject. In most cases, principals only nominated one such period. For these class periods, the principal was then asked to form class sections that were alike as possible in terms of student composition. Once this step was taken, MET researchers randomly selected a single class period and randomly assigned the pre-formed class sections to the teachers within a given exchange group. Thus, in Year One, specialist teachers had two sections in the study, whereas in Year Two, specialist teachers only had one section in the study. Teachers who could not be randomized continued in the study with one self-identified, naturally formed segment.

5.1.4.3 Non-Compliance with Randomization. Analyses of Year Two data by MET researchers showed substantial deviation from the plan to randomly assign student groups to teachers. These deviations from full randomization occurred for a number of reasons. One problem was that class groups were formed in the summer of 2010, before schools were certain which students or which teachers were going to appear when school opened. Following random assignment, some students transferred to other schools or to other teachers' classes in the same school; some teachers left teaching or taught different course sections or grades than planned. All of this led to non-compliance with the randomization regime. Moreover, in some cases, schools simply did not implement the randomization. The extent of this non-compliance is discussed in Section 5.2 below, where we discuss realized samples.

5.1.5 Recruitment of Students. The selection of teachers and their observed class sections determined the student sample for the study, and once students were identified, efforts were made to include ***<u>all</u>*** students from the classrooms selected for study. Students in these classrooms received informational fliers and consent forms to take home to their parents. In all districts but Hillsborough, a process of passive consent was used in which parents had the opportunity to remove their child from the study. In Hillsborough County Public Schools, students were required to bring in signed permission slips to be included as part of the study (i.e. active consent). Students in any district who opted out of participating did not take the student survey or supplemental assessments administered as part of the study, and during video recording of classroom instruction, they sat in a specific section of the room that was not video recorded. However, administrative data on student background and

state assessment scores for all students in a teacher's focal classes were obtained and used in the study.

5.2 Realized Samples

Table 3 shows the realized sample sizes for the MET *teacher* samples (for both years of the study). We focus on teacher samples in this section of the *User Guide* for several reasons. First, the primary unit of interest in the MET Study is teachers. Second, student and school samples were (in large part) determined by these teacher samples. As Table 3 shows, there was attrition in the number of teachers participating across both years of the study. The Year One sample (left hand column) shows all teachers who participated in Year One of the study, even if they ended up not being eligible for randomization in Year Two. The Year Two sample (right hand column) shows the number of teachers who participated in Year Two of the study broken down by their randomization status. Comparing the Year One sample to the Year Two sample, one can see a pattern of attrition in teacher sample size. In Year One of the study, a total of 2741 teachers in 317 schools took part in the MET Study, distributed across grade/subject groupings as indicated in the table. By contrast, the Year Two sample includes just 2086 teachers in 310 schools. Thus, as the table shows, about 24% of the Year One teacher sample was not included in the Year Two sample, with attrition across years varying by districts and ranging from about 21% of teachers in Denver to about 27% in Dallas.

There were a variety of reasons for teacher attrition. Some teachers (and their students) were lost when their school dropped out of the study (11 schools; 60 teachers). Additionally, individual teachers dropped out when they left their school or district, began teaching a different subject or grade, lost interest in the study, or became ill.

5.2.1 Realized samples versus national populations. The reader should understand that the MET Study used a process of "opportunity" sampling to recruit participants. In this process, the primary sampling units (districts) were selected as a matter of convenience by the MET Study, the schools within these districts were volunteers that met certain restrictions, the teachers within schools were volunteers, the classes chosen for the observational component were either self-identified by teachers (in Year One of the study and for non-randomized teachers in Year Two) or randomly assigned on the basis of scheduling constraints (in Year Two of the study), and the students in the MET sample were included in the study simply as a result of all these prior opportunistic processes.

With this in mind, the reader will see that the MET district sample is not representative (or even intended to be representative) of any identified universe of school districts. In fact, MET districts are among the largest school systems in the United States, ranging from the largest school district in the country (New York City) to the 63rd largest (Denver). Since the study was conducted in these large districts, the MET teacher sample is not a nationally representative sample of teachers. **Moreover, the MET LDB was not designed as a representative sample of teachers in each district. Consequently, comparisons across districts are not valid.** Analytical models should control for unobserved differences among districts, but conclusions about differences between districts cannot be derived from these data.

Table 3: Year One MET Teacher Sample *vs*. MET LDB Core Teacher Sample by Focal Grade/Subject		
	Full Sample All Year One Teachers (AY 2009-2010)	**Core Study Sample All Teachers Present in Year Two (AY 2010-2011)**
4th and 5th Grade English/Language Arts (ELA)	**138**	**Randomized: 98** **Non-Randomized: 29**
4th and 5th Grade Mathematics	**102**	**Randomized: 67** **Non-Randomized: 31**
4th and 5th Grade ELA and Mathematics	**634**	**Randomized: 305** **Non-Randomized: 52**
Grades 6-8 ELA	**606**	**Randomized: 292** **Non-Randomized: 139**
Grades 6-8 Mathematics	**528**	**Randomized: 282** **Non-Randomized: 120**
Grades 6-8 ELA and Mathematics	**18**	**Randomized: 4** **Non-Randomized: 4**
9th Grade Algebra I	**233**	**Randomized: 116** **Non-Randomized: 44**
9th Grade English	**242**	**Randomized: 108** **Non-Randomized: 48**
9th Grade Biology	**240**	**Randomized: 103** **Non-Randomized: 60**

5.2.2. Realized teacher samples versus district populations. An interesting question is whether the teachers who volunteered for the MET Study differed from the population of teachers in the districts where they were employed. Partial data on this point were presented in the MET research report entitled *Gathering Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains* (Gates Foundation, 2012: Table 2, page 17; Table 10, page 17) [8]. These data suggest that in Year One of the study, teachers who participated in the MET Study were similar to teachers in the same districts who did *not* participate in the MET Study (across a range of characteristics, including ethnicity, years of teaching experience, and a "value-added" measure of teaching effectiveness based on state assessments). We conducted an additional analysis to compare the Year Two teacher sample to the Year One teacher sample. That analysis found that, despite substantial attrition in the teacher sample, the Year One and Year Two MET teacher samples were alike across a range of variables, including ethnicity, and years of teaching experience.[9]

5.2.3 Non-compliance rates for randomization. As discussed earlier, MET researchers' attempts to randomly assign classes to teachers in Year Two of the study was subject to non-compliance. This is not unusual, for some amount of non-compliance is to be expected in most random assignment

[8] The cited report only covers teachers in grades 4-8 who were randomized in year two.

[9] The interested reader can also consult the paper by Kane, McCaffrey, Miller, and Staiger (2012) entitled *Can We Identify Effective Teachers? Validating Measures of Effective Teaching using Random Assignment* (available at www.metproject.org) to compare samples across years and to district-wide populations. Table 2 of that paper shows that the MET year two teacher and student samples were quite similar to the Year One MET samples of teachers and students *and* to the populations of teachers and students in participating districts.

experiments involving human research subjects. One reason non-compliance occurred in the MET Study was that random assignments of classes to teachers occurred in the summer of 2010, before schools were certain which students or teachers would be present when school opened. Following random assignment, some students transferred to other schools or to other teachers' classes in the same school and some teachers left teaching or taught different course sections or grades than planned. In other cases, schools simply did not implement the randomization, as demonstrated by the finding that students randomly assigned to one teacher ended up together, but in another teacher's classroom. MET researchers could not prevent these failures in compliance.

Table 4 shows the extent of non-compliance by district for the 4th through 8th grade samples and the pooled non-compliance rate for the high school sample. The table shows the percentage of students in a district who: (a) remained in the classroom to which they were randomly assigned; (b) moved to another classroom in an "exchange" group (*aka*, randomization block); (c) remained in the school but were in a classroom outside the exchange group; (d) moved to another school within the district; or (c) were missing. The reader will note that non-compliance was extensive in all districts. Readers interested in an additional discussion of non-compliance to randomization, resulting teacher and student samples, and procedures for analyzing data in light of non-compliance should refer to the paper by Kane, McCaffrey, Miller, and Staiger (2012), entitled *Can We Identify Effective Teachers? Validating Measures of Effective Teaching using Random Assignment*. This paper is available at the MET Study web site (www.metproject.org).

Table 4: Percentage of Students Randomized in MET Study

Proportions	Remaining in Randomization Block		Remaining in School	Other Schools in District	Missing	Total
	Same Teacher	Different Teacher				
4th to 8th Grade Sample by District (Math & ELA Stacked):						
Dallas	0.65	0.02	0.25	0.01	0.05	1.00
Charlotte Mecklenburg	0.62	0.14	0.18	0.02	0.04	1.00
Hillsborough	0.55	0.16	0.23	0.03	0.03	1.00
New York	0.44	0.20	0.21	0.01	0.14	1.00
Denver	0.40	0.24	0.24	0.00	0.12	1.00
Memphis	0.30	0.22	0.32	0.01	0.15	1.00
High School Sample:						
Math, ELA and Biology	0.45	0.14	0.30	0.04	0.07	1.00

6.0 ADDITIONAL INFORMATION ON DATA COLLECTION, INSTRUMENTS AND MEASURES

This section of the *User Guide* turns from issues of sampling to issues of data collection, instrumentation, and measurement. The section focuses (a) value-added measures of student learning; (b) the student perceptions (or TRIPOD) survey; (c) classroom observation protocols; (d); the content knowledge for teaching assessment; (e) the teacher working conditions survey; (f) the teacher survey; and (g) the MET principal survey. The discussion is brief, but readers needing more information can consult the MET Study web site (www.metproject.org).

6.1 Value-Added Measures of Student Learning

The MET Study used a variety of student achievement assessments to examine teacher effects on student learning. One set of measures came from the state assessments administered in each district. In general, these were multiple choice tests, and a worry in the education research community has been that such state assessments often measure only basic skills. For this reason, MET researchers also administered a set of supplemental assessments as part of the study. The supplemental assessments were chosen because they consist of more cognitively demanding test content and presented students with constructed response items. The reader should note that state assessments were administered according to state-specific timelines and procedures and were thus administered to all eligible students, while administration of supplemental assessments was overseen by Westat (a large research firm), conducted under timelines imposed by the MET study using procedures recommended by the testing firms that published the tests, and conducted only with consenting students. The reader also will note slight variation (across tests and districts) in testing dates, both for state assessments and supplemental assessments. Tables 5a and 5b show testing dates for both years of the study, by district. Over 93% of students have state test scores from the year they were in the study. Around 78% of students have state test scores from the year before they were in the study.

6.1.1 SAT-9 open-ended (OE) reading assessment. One of the supplemental assessments administered in all ELA classrooms in grades 4-8 was the Stanford 9 OE Reading Assessment. This assessment was administered in a single period. The assessment presented students with one extended reading passage and then asked them to respond to nine, open-ended tasks (which required students to provide short, written responses to comprehension questions). In the MET Study, all SAT-9 OE passages consisted of narrative text. The publisher reports a developmental scale score for the SAT-9 open-ended reading assessment so that student scores can be compared across grades. However, MET researchers did not administer the standard form and thus do not report these scale scores. During Year One, 75% of students in 4th-8th grade ELA sections completed the SAT9. During Year Two, 80% completed the SAT9.

6.1.2 Balanced assessment of mathematics (BAM). The Balanced Assessment in Mathematics was administered in all Mathematics classrooms at grades 4-8. BAM is an open ended assessment that seeks to measure students' capacities in the following areas: modeling/formulating problems;

Table 5a: Student Testing Windows in Year One (by District)				
District	**Stanford 9 ELA**	**BAM**	**ACT Quality Core**	**State Assessments**
Charlotte	**May 21-June 7**	**May 21-June 7**	**May 21-June 7**	**March 9-20; April 22-May 14; May 3-7; May 25-June 10**
Dallas	**May 25-27**	**May 25-27**	**No High Schools**	**March 22-April 2; April 26-30; May 10-14; May 19-26**
Denver	**April 19-May 7**	**April 19-May 7**	**April 19-May 7**	**March 1-19; April 26-May 6**
Hillsborough	**May 3-14**	**May 3-14**	**May 3-14**	**March 9-19; March 29-April 29; April 19 – May 19**
Memphis	**May 3-7**	**May 3-7**	**May 10-14**	**April 12-16; April 29; May 11-12; May 19-21**
New York City	**May 24-June 4**	**May 24-June 4**	**May 31 –June 11**	**April 26-28; May 5-7; June 14-24**

Table 5b: Student Testing Windows in Year Two (by District): Supplemental Assessments Only*			
District	**Stanford 9 ELA**	**BAM**	**ACT Quality Core**
Charlotte	**April 27 –May 6**	**April 27 –May 6**	**Dec 13-21; April 27- May 13; May 16-20**
Dallas	**May 10-20**	**May 10-20**	**No High Schools**
Denver	**April 11-21**	**April 11-21**	**April 11-21**
Hillsborough	**May 2-13**	**May 2-13**	**April 18-21; May 2-5**
Memphis	**April 25-May 6**	**April 25-May 6**	**May 2-13**
New York City	**May 16-27**	**May 16-27**	**May 16-27**
***States assessment schedules approximate those shown in Table 3a**			

transforming/manipulating mathematical formalisms; inferring/drawing conclusions; and communicating about mathematics. Each assessment form contains four to five assessment tasks and requires 50-60 minutes to complete. In the MET Study, concerns about generalizability led the MET researchers to administer three different forms of the BAM (from the relevant grade levels tests for 2003, 2004 and 2005) in each classroom. Thus, all students in a classroom (or class section) did not take the same test. Assessment tasks are scored on a four-point scale (ranging from attribute not present to attribute predominantly present) for each of the dimensions of thinking being assessed on the BAM (i.e., modeling/formulating; transforming/manipulating; inferring/drawing conclusions; and communicating). MET data files include a student's scores on each of these dimensions for each task, as well as the test form administered to that student. During Year One, 79% of students in 4th-8th grade math classes completed the BAM. During Year Two, 81% completed the BAM.

6.1.3 ACT QualityCore. ACT QualityCore Assessments were the only student achievement tests administered to students in the MET Study at grade 9. The assessments were English-9, Algebra I, and Biology, with students taking the particular assessment corresponding to the subject of their MET section. In the MET Study, the end of course assessments include one form (administered during a class period) that included multiple choice items only, and another form (administered during a class period) that included 1-3 constructed-response items. ACT Quality Core Assessment scores are reported as a developmental scale score that can be used to track students' growth across multiple courses in the same subject area. During Year One, 66% of 9th grade students had valid scores on an ACT assessment. This response rate ranged from 57% in math to 61% in Biology to 72% in ELA.

During Year Two, 76% of students had valid scores on an ACT assessment (70% math, 75% Biology, 80% ELA).

6.1.4 Construction of value-added measures. MET researchers used student test scores to construct "value-added" measures of teaching effectiveness for individual teachers. This process began with the implementation of a student rostering process in MET classrooms. In this process, lists of students in a MET classroom were presented to teachers each time student data were collected, and teachers were given the opportunity to inspect these lists, provide any updates or deletions (along with the dates at which any listed students exited or new students entered), and to indicate the percentage of instructional responsibility they had for each listed student. A final roster was presented to teachers at the end of the school year to verify that all data were correct.

Using these rosters, the second step in creating value-added measures was to connect roster data to student achievement data and then estimate value-added measures of teaching effectiveness. It is important to note that the statistical models used to calculate a value-added score for teachers in the MET Study were based on a single outcome measure (e.g., the state ELA test, the state Mathematics test, the SAT-OE, or the BAM). In addition, MET researchers estimated value-added models for each outcome separately by grade and district, so that **all value-added estimates are not just test-specific, but also grade and district specific**. With the data sub-divided in this way, MET researchers estimated the following equation for each test outcome, at each grade, in each district:

$$S_{it} = X_{it}\beta + \bar{X}_{jkt}\gamma + \theta Z_{i(t-1)} + \lambda \bar{Z}_{jk(t-1)} + \varepsilon_{it} \qquad (1)$$

where: (a) S_{it} is one of the achievement outcomes for student i in the current school year t; (b) X_{it} is a vector of student background variables for student i at time t (described in footnote 10); (c) $\overline{X}_{jkt}$ was a vector of the same student background variables, only this time measured as averages for the section k taught by teacher j in which the student was enrolled at time t; (d) $Z_{i(t-1)}$ is student i's prior test score (in the same subject) at year t-1; (e) $\overline{Z}_{jk(t-1)}$ is the mean prior year test score of all students in the section k taught by teacher j in which student i is enrolled; and (f) ε_{it} is a student-specific residual, characterized by the usual regression assumptions.[10]

After estimating this model, the residuals (ε_{it}) were used as the basis for constructing MET teachers' value-added scores. For elementary level generalist teachers, the value-added score for a teacher is simply the average of residuals across all the students in a teacher's class; for specialist teachers, the value-added score averages residuals to the section level (i.e., to a given section k). Standard errors

[10] The reader should note that all student achievement scores were first converted to rank-based z-scores (by ranking students and then taking the inverse normal distribution of the ranked scores). State tests were ranked within district, subject, and grade. The SAT9-OE and ACT tests were ranked across districts, but within subject and grade. The BAM test was ranked across districts, but within grade and test form.

For state-test outcomes, the prior achievement variable is the rank-based z-score of the achievement on the prior year's state test in the same subject. For SAT-9 OE, BAM, and ACT, the prior achievement variable is the rank-based z-score of the prior year's state test in the same subject.

The reader should note that the student-level covariates used in these regressions differed across districts. In Charlotte-Mecklenburg the covariates were: ethnicity, ELL status, age, gender, special education, gifted status; in Dallas they were: ethnicity, ELL, age, gender, special education, free or reduced lunch; in Denver they were: ethnicity, age, ELL, free or reduced lunch, gender, and gifted; in Hillsborough they were: ethnicity, ELL, age, special education, gifted status, and free or reduced lunch; in Memphis they were: ethnicity, ELL, free or reduced lunch, gender, gifted, special education; in NYC they were: ethnicity, ELL, gender, special education, free or reduced lunch.

for each teacher's VAM are calculated as the standard deviation of the residuals for a class (k) divided by the square root of the number of student residuals included in the classroom average.

6.2 The Student Perceptions (TRIPOD) Survey

The Student Perception (aka Tripod) Survey was administered to all consenting students in class sections taught by MET teachers. The survey is designed to measure seven dimensions of classroom instruction referred to by instrument developers as the "Seven Cs" (7Cs). These dimensions are: Care, Control, Clarify, Challenge, Captivate, Confer, and Consolidate. Care measures students' perceptions of whether the classroom is a safe place. Clarify measures students' perceptions of teacher behaviors that help students' to better understand the content being taught. Challenge measures students' perceptions of classroom rigor and required effort. Captivate measures students' perceptions of how well the teacher captures the attention and interest of students. Confer measures students' perceptions of how much a teacher takes students' points of view into account when teaching. Consolidate measures students' perceptions of how much the teacher helps students cognitively represent what they have learned in a connected way and how well the teacher promotes students' understanding of the interconnectedness of different curriculum topics.

There were scales in addition to the 7Cs on the student perception survey. In both years, students report on effort they exert in class, how happy they are in class, and the amount of test prep activities they engage in. During Year One only, students reported on their college aspirations and how often they read at home. During Year Two only, the student perception survey contained Carol Dweck's implicit theory of intelligence scale that measures the degree students believe effort versus inherent ability lead to success. The Year Two survey also contained Angela Duckworth's measure of academic grit, which measures students' willingness to persist on cognitively challenging problems. As discussed earlier, two versions of the survey were administered, one for students in grades 4-5, and one for students in grades 6-9. In classes for generalist teachers, a randomly selected half of the class filled out the survey while thinking about their ELA class and the other half completed the survey while thinking about the Math class. Most questions on a Tripod survey use Likert-type response options with a 5-point scale (Totally Untrue to Totally True). Scales were created by taking the simple mean of the items composing the scale. MET researchers tended to aggregate TRIPOD survey results to the class level using simple means and then adjust these aggregates for classroom level student characteristics[11].

6.3 Classroom Videos and Video Scoring Processes

The MET Study used a variety of observation protocols to measure the quality of classroom teaching. This section of the *User Guide* describes how the classroom videos (that were scored with these protocols) were collected, how video scoring was conducted, and the observation protocols that were used in the study and the scores derived from these protocols.

6.3.1 Video data collection. Video recording occurred between February and June of 2010 in Year One of the study and between October 2010 and June 2011 in Year Two of the study. The recording of

[11] In order to adjust the classroom level aggregates, MET researchers regressed the aggregates on the same classroom-and student-level student characteristics as were used in the VAM models (see section 6.1.5). This was done separately for each district and grade. The residuals from this regression form the adjusted classroom level student perception survey scores. These scores are included in the classroom level file described below.

each teacher was spread out within these periods in an attempt to assure that lessons were more representative of instruction than would be recording within a very compressed time period. In addition, a goal of MET researchers was to video teachers as they taught different curricular topics. To achieve this goal, MET researchers asked teachers to schedule half of their days of video recording when they were teaching a set of "focal" topics determined by the researchers, and to schedule the other half when they were teaching a topic of their choice. Thus, generalist teachers were recorded teaching a focal topic for two days and teaching a topic of their choice for two days. Each day recorded both the ELA and the Mathematics lesson. Specialist teachers were recorded teaching a focal topic on one day (in both MET sections) and teaching a topic of their choice on one day (in both MET sections). In the Core Files the focal topic (if any) being taught by a teacher is listed as a variable.

Teachers were trained and were responsible for all video recording as well as for uploading video to a secure website. A specially-designed camera rig was used for the recording. Each rig had two cameras: one focused on the board, the other providing a 360 degree classroom view. The rig also included two wireless microphones, one to capture the teacher's voice, the other to capture student voices. Captured videos were uploaded, and a research partner combined the separate video and audio channels into one video. This video was then made available to teachers who were required to check the video for accuracy and upload student work, student assignments, lesson plans, and written commentaries on the lesson.[12]

6.3.2 Raters and rater training. Videos were scored by 902 current and former teachers trained by MET researchers. Training for FFT and PLATO occurred through a web interface developed by MET researchers; training for CLASS and MQI occurred through websites already established by protocol developers[13]. Training was always self-directed and lasted between 17 and 25 hours depending on the instrument. Training included four main sections: training in the web interface used to access videos; training on how to eliminate bias in scoring; frame of reference training that provided an overview of the protocol; and specific training on the scoring of each scale in a protocol. This last area of training required raters to watch and score videos and was followed by practice sessions and an initial certification test.

6.3.3 Phases of scoring. Scoring of videos by raters proceeded in a number of phases. (See below Section 8.0 Comparison of Phase 1 and Phase 2 Video Scoring). The scoring process differed across these phases, and it included different samples of teachers. As a result, the Core Files include data that was scored during Phase 2 of the scoring process (using all of the videos from Year One teachers who also were randomized in Year Two of the study).

6.3.4 Assignment of raters to videos. In Phases 2 and 3 of the scoring process, raters were assigned to videos by the following process. First, videos were grouped so that the same group of raters scored all videos within an "exchange group" (or randomization block) of teachers. Then, restrictions were placed on which raters could score particular videos. Under these rules, raters were not allowed to score videos of teachers they knew; raters were not allowed to score videos of teachers in a district where they had worked or had other affiliation; and raters were only allowed to score a single video

[12] The process was slightly different for biology classes doing labs. When biology labs were filmed, handheld video rigs were used. These handheld rigs combined an IPhone, an IPod, and wireless microphones to collect different views of the classroom.

[13] The UTOP sub-study used a different scoring procedure. As UTOP was only used on a small sub-set of videos and followed different procedures, we consider it a sub-study rather than part of the main study.

for a given teacher in a given year. Once these restrictions were in place, raters were randomly assigned to groups of teachers and scored one video of each teacher in a block[14].

6.3.5 Scoring procedures. During Phases 2 and 3 of the scoring process, raters scored videos during regular, four-hour shifts, during which time they scored only the first 30-35 minutes of each video assigned to them. Scoring was generally done in time segments. For example (as discussed in more detail below), videos scored with the CLASS were divided into 15-minute segments, whereas videos scored with MQI were divided into 7.5 minute segments. Raters recorded scores on each item in an observation protocol for each of these segments.

Each rater was assigned to score in four-hour shifts, during which time the rater scored all assigned videos using only *one* observation protocol. Videos were scored using a web-based interface that eliminated the need for raters to gather at a central location. The online system ensured that raters were shown and coded the correct segments. In scoring, raters were only allowed to pause a video in order to record comments, and, after watching a segment in its entirety once, raters also were allowed to rewind and fast forward a video to re-watch important parts.

The MET Study used a variety of procedures to assure that raters were using observation protocols faithfully. Each rater's shift began with the scoring of calibration videos (i.e., videos that had been assigned "true" scores), and raters had to score these calibration videos at a pre-established level before being allowed to score their assigned videos. Raters also scored validity videos. These videos have assigned "true" scores and were mixed in with their daily scoring assignment (5% of videos scored by a rater were validity videos). Raters did not know when they were scoring a validity video. These videos provide another view of the accuracy of scoring. In addition, during the scoring process, raters had full access to training materials and training videos that contained true scores. Finally, the work of each rater was closely monitored by a scoring leader, and if raters ran into problems scoring a video, they were able to send the video to the scoring leader to score or request a joint review session with that leader. In addition, scoring leaders "back-scored" at least one video from each rater during each shift to assure that raters were scoring videos accurately and/or identify raters with low accuracy who might need additional mentoring. In addition, scoring leaders were expected to contact each rater at least once during a shift in order to check on progress and provide useful feedback, usually in the form of a joint review or conversations about back-scored segments. Finally, scoring leaders had the authority to suspend a rater's scoring privileges, to force raters to redo the daily calibration, or to terminate a shift if a rater's scoring became problematic.

6.3.6 Double-scored videos. About 5% of Phase 2 and Phase 3 videos were scored by two raters[15]. Double-scored videos were selected at random, and raters did not know that they were scoring a video that had previously been scored. MET researchers do not specifically cite data on inter-rater reliabilities using common indices of this statistic, such as Cohen's Kappa. However, they do provide data showing that for most of the observation protocols, about 10 percent or less of the total variance in scores was due to "main" rater effects—that is, effects that arise because some raters consistently score high and other raters consistently score low. The main exceptions were MQI (particularly for in the domains of "errors and imprecision" and "explicitness") and to a lesser extent PLATO (particularly for "modeling" and "strategy use and instruction"). These data are reported in *Gathering*

[14] For CLASS only, raters were assigned to score a segment instead of a video. Thus, CLASS raters scored one 15 minute segment for each video in the exchange group. On all other instruments, the raters scored an entire video for each teacher.

[15] For QST only, 10% of videos were double scored.

Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains (Gates Foundation, 2012: 35-36).

6.4 Classroom Observation Instruments and Derived Scores

Table 6 lists the classroom observation protocols used in the MET Study and the samples of videos scored using each protocol. The remainder of this section describes these protocols in more detail. Additional detail on these instruments is reported in *Gathering Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains* (Gates Foundation, 2012: 18-19), on the MET Study web site, and at the ICPSR MET LDB web site. Instrument abbreviations (shown in Table 6) are explained in the discussion to be presented.

6.4.1 The Classroom Assessment Scoring System (CLASS). CLASS is an observational protocol designed to measure the extent to which teachers effectively support children's social and academic development. Two different versions of CLASS were used in the MET Study: the Upper Elementary (Grades 4-5) and the Secondary (Grades 6-9).

The CLASS instrument is divided into three broad domains of measurement: Emotional Support, Classroom Organization, and Instructional Support. Each domain, in turn, is measured by a number of dimensions. The domain "Emotional Support," for example, refers to the emotional tone in a classroom, which can be measured along four dimensions: positive climate, negative climate, teacher sensitivity, and regard for student perspectives. The domain "Classroom Organization" refers to the ways a classroom is structured to manage students' behavior, time, and attention, which can be measured along three dimensions: behavior management, productivity, and instructional learning formats. The domain "Instructional Supports" refers to the ways a teacher provides supports to encourage student conceptual understanding and student problem solving and can be measured along four dimensions: content understanding, analysis and problem solving, instructional dialogue, and quality of feedback. CLASS also includes a fourth domain called "student engagement," which includes only a single scoring dimension.

Table 6: Classroom Observation Protocols Used in the MET Study and Included in Core Files

Instrument	Target Videos	Videos Scored
CLASS	**This is a "generic" instrument that can be used to score the teaching of any subject. Two forms were used: the upper elementary form and the secondary form. In the MET Study, the goal was to score ELA and Math videos from grades 4-5 using the upper elementary form, and ELA, Math, and Biology videos from grades 6-9 using the secondary form.**	***All* videos of randomized teachers except for Biology videos (plus Phase 1 videos).**
FFT	**This is a "generic" instrument that can be used to score the teaching of any subject. FFT includes 4 domains, but the domains of planning and preparing and professional responsibilities were not coded. In the MET Study, this protocol was used to score ELA, Math, and Biology videos from grades 4-9.**	***All* videos of randomized teachers except for Biology videos (plus Phase 1 videos)**
MQI "Lite"	**This is a "subject-specific" instrument designed for observations of mathematics teaching. The protocol used in the MET Study was modified to include only 6 of the 24 original elements in the protocol. In the MET Study, this protocol was used to score Math videos from grades 4-9.**	***All* videos of randomized Math teachers (plus Phase 1 Math videos)**

Table 6: Classroom Observation Protocols Used in the MET Study and Included in Core Files		
Instrument	**Target Videos**	**Videos Scored**
PLATO Prime	**This is a "subject-specific" instrument designed for observations of ELA teaching. The protocol used in the MET Study (known as PLATO "Prime") was modified to include only 7 of the original 13 original scoring elements. In the MET Study, this protocol was used to score ELA videos from grades 4-9.**	***All* videos of randomized ELA teachers (plus Phase 1 ELA videos)**
QST	**This is a "subject-specific" instrument designed to score Biology teaching. In the MET Study, this protocol was used to score grade 9 Biology videos.**	***All* videos of randomized Biology teachers**

CLASS scoring is done using a detailed scoring rubric. In this rubric, a classroom is scored on each instructional dimension at 15-minute intervals using a 7-point scale. For the MET Study, only the first 30 minutes of each video was scored. Scores are assigned based on anchor descriptions of what is to be observed in order for a classroom to be scored at "high," "mid," and "low" points on the 7-point scale. In the MET Study, dimension scores are often aggregated to higher levels of analysis simply by averaging raters' scores to get a single segment score and then calculating the harmonic mean of segment scores across all segments for a particular target of measurement (e.g., a day, a class section, a teacher). Standard errors of measurement for these derived scores are not generally reported.

6.4.2 Framework for Teaching (FFT). FFT is an observational protocol designed to measure components of instruction that reflect a constructivist view of teaching. The instrument was used in the MET Study to code instruction in all English and math classrooms at grades 4-9.

The FFT instrument divides the complex activity of teaching into four domains of teaching responsibility: Planning and Preparation (Domain 1), Classroom Environment (Domain 2), Instruction (Domain 3), and Professional Responsibilities (Domain 4). However, as discussed earlier, the MET
Study scored videos on only two of these domains ("Classroom Environment" and "Instruction"). Each of these domains, in turn, is measured by a number of dimensions. The domain "Classroom Environment," for example, is measured along five dimensions: creating an environment of respect and rapport; establishing a culture for learning; managing classroom procedures; managing student behavior; and organizing physical space. The domain "Instruction" also is measured along five dimensions: communicating with students; using questioning and discussion techniques, engaging students in learning, using assessment in instruction, and demonstrating flexibility and responsiveness.

FFT scoring is done using a detailed scoring rubric. However, unlike the other observation protocols used in the MET Study, raters use this rubric only once for each video, that is, a video scored with the FFT protocol will have only 1 score per video on each of the dimensions just described. In the MET Study, raters did not watch an entire video, however. Instead, raters watched for 15 minutes at the beginning of the video and then ten additional minutes at the 25-35 minute mark. They then scored the video as a single segment[16]. Each dimension of the FFT protocol was scored on a 4-point scale ranging from unsatisfactory, to basic, to proficient, to distinguished, and for each dimension a

[16] During Phase 3 scoring, which focused on videos shorter than 25 minutes, raters watched the video in its entirety before scoring the whole video as a segment.

detailed scoring rubric described the kind of evidence appropriate for assigning a particular score. In the MET Study, scores for a dimension are often aggregated to higher levels of analysis simply by averaging raters' scores to get a single video score and then calculating the harmonic mean across videos for a particular target of measurement (e.g., a class section, a teacher). Standard errors of measurement for these derived scores are not generally reported.

6.4.3 Mathematical Quality of Instruction (MQI lite). MQI is an observational instrument designed to measure the mathematical work that occurs in classrooms, on the theory that such work is distinct from classroom climate, pedagogical style, or the deployment of generic instructional strategies. The MQI was used in the MET Study to code instruction in Mathematics at grades 4-9.

The MQI instrument measures the mathematical quality of instruction by assessing classroom instruction along six dimensions: Richness of the Mathematics; Errors and Imprecision; Working with Students and Mathematics; Student Participation in Meaning-Making and Reasoning; Explicitness and Thoroughness; and Connections between Classroom Work and Mathematics. The dimension Richness of Mathematics captures student meaning making and classroom mathematical practices. The dimension Errors and Imprecision captures major errors made by the teacher, imprecision in language and notation used by the teacher, and lack of clarity. The dimension Working with Students and Mathematics captures how the teacher responds to students' mathematical ideas and remediates student errors. The dimension Student Participation in Meaning-Making and Reasoning captures how students ask questions and reason about mathematics, students provide mathematical explanations, and the cognitive requirements of tasks. The dimension Explicitness and Thoroughness captures how complete the mathematical ideas and concepts are during the lesson. The dimension Connections between Classroom Work and Mathematics captures whether classroom activities are aimed at developing mathematical ideas. Raters score each segment on these 5 dimensions as well as giving an overall video score for each dimension. In addition to the main dimensions of the MQI, scorers rate a teacher on his or her apparent mathematical knowledge for teaching and provide a holistic score for the quality of the entire video.

MQI segment scoring was done by dividing the first 30 minutes of each video into 7.5 minute segments and then scoring each segment on the dimensions just described. The scoring rubric is built around a 3-point scale (dimension not present, partially present, or predominantly present), and scoring anchors are provided for each score on each dimension. Once all segments are scored, raters also assign an overall score to the video. The overall score is based on a 3-point scale ranging from low to medium to high. In the MET Study, scores for each dimension are often aggregated to higher levels of analysis simply by averaging raters' scores to get a single segment score and then calculating the harmonic mean across segments for a particular target of measurement (e.g., a day, a class section, a teacher). Standard errors of measurement for these derived scores are not generally reported.

6.4.4. Protocol for Language Arts Teaching Observation (PLATO Prime). PLATO is an observational instrument designed to measure four underlying qualities of ELA instruction: disciplinary demand of classroom activity and discourse; instructional scaffolding of ELA content; representations and use of content; and the classroom environment. PLATO was used in the MET Study to code instruction in ELA at grades 4-9.

As used in the MET Study, seven (of 13 possible) dimensions of teaching were scored: Intellectual Challenge; Representation of Content; Models/Modeling; Explicit Strategy Instruction; Classroom Discourse; Behavior Management; and Time Management. In addition to these dimensions, raters also indicated whether the segment included instruction in the following content domains: Reading,

Writing, Literature, Speaking and Listening, Word Study, Grammar/Spelling, and Research Strategies. PLATO segment scoring was done by dividing the first 30 minutes of each video into 15 minute segments and then scoring each segment on the dimensions just described. For most dimensions, the scoring rubric is built around a 4-point scale (from Almost No Evidence to Limited Evidence to Evidence with Some Weaknesses to Consistent Strong Evidence), and scoring anchors are provided for each score on each dimension. The Representation of Content dimension was scored either 0 or 1. In the MET Study, scores for each dimension are often aggregated to higher levels of analysis simply by averaging raters' scores to get a single segment score and then calculating the harmonic mean across segments for a particular target of measurement (e.g., a day, a class section, a teacher). Standard errors of measurement for these derived scores are not generally reported.

6.4.5 Quality of Science Teaching (QST). QST is an observational instrument for science teaching that was developed as part of the MET Study. QST was used in the MET Study to score Biology instruction at grade 9. It was the only instrument used to score Biology videos.

QST measures three domains of science teaching: Assessing Teacher Content Knowledge; Engaging All Students in Learning Science; and Promoting Laboratory-based Inquiry. Each domain has three to four dimensions that are independently scored. The domain "Assessing Teacher Content Knowledge" includes three dimensions: Sets the Context and Focuses Learning on Key Science Concepts; Uses Representations; Demonstrates Content Knowledge; and Provides Feedback for Learning. The domain "Engaging All Students in Learning Science" includes three dimensions: Promotes Students' Interest and Motivation to Learn Science; Assigns Tasks to Promote Learning and Addresses the Demands of the Task for All Students; Uses Modes of Teaching Science Concepts; and Elicits Evidence of Students' Knowledge and Conceptual Understanding. The domain "Promoting Laboratory-based Inquiry" includes the dimensions: Initiates the Investigation; Provides Guidelines for Conducting the Investigation and Gathering Data; and Guides Analysis and Interpretation of Data.

QST scoring was done differently from most other instruments[17]. The instrument was broken into three "Groups of Scales" that were scored by separate raters. The first Group of Scales consists of the domain Assessing Teacher Content Knowledge. The second Group of Scales consists of the domain Engaging All Students in Learning Science. The third Group of Scales includes the domain Promoting Laboratory-based Inquiry as well as the dimension Elicits Evidence of Students' Knowledge and Conceptual Understanding from the domain Engaging All Students in Learning Science. Both the second and third Group of Scales code the dimension Elicits Evidence of Students' Knowledge and Conceptual Understanding.

QST scoring was done differently depending on the content of the video as reported by the teacher. Videos of laboratory experiments were only scored with the third Group of Scales. Videos not containing laboratory experiments are scored with the first two Groups of Scales. When using the first two Groups of Scales, the first 30 minutes of a video were scored as two 15 minute segments. Scores were given on a 4 point scale (ranging from Low to Low Medium to High Medium to High). When using the third Group of Scales, the first 60 minutes of a video were scored as up to four 15 minute segments. Scores were given on a 5 point scale (ranging from 0-4). In the MET Study, scores for each dimension are often aggregated to higher levels of analysis simply by averaging scores by different raters to get a single segment score and then calculating the harmonic mean across

[17] QST was not scored in Phase 1. However, a scoring method similar to the methods used to score other instruments during Phase 1 scoring was used to score the QST protocol.

segments for a particular target of measurement (e.g., a day, a class section, a teacher). Standard errors of measurement for these derived scores are not generally reported.

6.5 Content Knowledge for Teaching Assessment (CKT)

MET researchers administered several forms of an assessment designed to measure teachers' content knowledge for teaching. These assessments were specifically designed for and administered to: (a) ELA teachers in grades 4 through 6; (b) ELA teachers in grades 7 through 9; (c) Mathematics teachers in grades 4 and 5; (d) Mathematics teachers in grades 6 through 8; and (e) Algebra I teachers in grade 9. The assessment for teachers of Mathematics in Grades 4 and 5 was administered to participating MET teachers in the fall of 2010 and the other four assessments were administered in early 2011, with the testing window ending in April.

The ELA assessments attempted to measure knowledge for teaching closely tied to the teaching of ELA, such as: choosing a text to support a specific teaching goal; selecting an activity to highlight a particular feature of a text or literary technique; choosing an activity to assess students' understanding; and analyzing student writing for weaknesses or strengths. The Mathematics assessments attempted to measure knowledge for teaching tied to the teaching of Mathematics, such as: choosing and using appropriate mathematical representations; choosing examples to illustrate a mathematical concept; interpreting student work, including use of nonstandard strategies; and evaluating student understanding.

Each assessment form included two types of selected-response items: single-selection multiple-choice items and multiple-response table questions. The table questions were each composed of three to six items. In addition, each of the assessments (except for Algebra I) included two constructed-response (open-ended) questions. Scores reported in the MET data files are based on the number of correct selected-response items on the assessment combined with a total score for constructed-response items. Constructed-response items were scaled on a 3-point scale. The overall scale score that is reported is simply a linear transformation of the total score to give a possible range of 0-100.

6.6 Teacher Working Conditions Survey (Year One)

MET researchers administered a survey to all teachers at MET schools during Year One[18]. The survey asked teachers to report on levels of support in their school environment. Specifically, teachers reported the quality of the school facilities and availability of resources to support instruction; the extent that schools protect their time to plan and provide effective instruction; opportunities for professional development and the quality of that professional development; support to help teachers analyze student data and collaborate to improve instruction; support given to teachers in managing student behavior; the degree of teacher leadership present at the school; trust of school leadership and the level of support received from school leadership; and the level of parent and community involvement.

Surveys were administered through a confidential online system. Because of the sensitive nature of the information collected, teacher responses were only linked to the teacher's school. There was no

[18] Because of pre-existing contracts, the survey was administered to all teachers in the district in Charlotte-Mecklenburg, Denver, and Memphis.

direct link between a survey response and a specific teacher. However, teachers were able to self-identify at MET Study sites and provide their MET ID.

6.7 Teacher Survey (Year Two)

During Year Two of the study, a different teacher survey was administered. This survey was administered only to MET Study participants. It asked teachers to report very broadly on current evaluation practices in their school, including trust in their principal; the credence given to ratings provided by administrators; the receipt of feedback on practice as well as whether action based on that feedback was taken and the result of such actions. The survey was administered in the early summer immediately following Year Two of the study. Teachers filled out the survey confidentially through a web based portal.

6.8 Principal Survey

A survey was administered to all MET school principals (approximately 330) just after the 2010-2011 school year (Year Two) ended. The survey contained items asking principals about their school's current teacher evaluation policies, trainings they received on teacher evaluation, comfort with teacher evaluation, and their perceptions of the effectiveness of the teacher evaluation system currently in place in their district. In addition, school principals were asked to rate the effectiveness of up to 12 participating MET teachers in their school using a six-point scale that ranged from Exceptional (top 5%) to Very Good (top 25%) to Good (top 50%) to Fair (top 75%) to Poor (bottom 25%) to Very Poor (bottom 5%). Additionally, principals could indicate that they were unable to rate their teachers. Principals gave a confidence rating to their judgments of teacher effectiveness on a 5 point scale ranging from Strongly Confident to Not at All Confident. Last, each principal reported how many formal observations and how many informal observations he or she had conducted with each teacher.

7.0 STRUCTURE OF THE CORE FILES (ICPSR 34414)

As discussed in previous sections of this *User Guide*, MET researchers gathered many different kinds of data on participating school districts, schools, teachers, class sections, and students during the course of the MET Study. The MET LDB Core Files have been constructed to facilitate use of the MET data holdings by a broad user group. The Core Files are designed to provide the most complete data consistent with the MET study design as described above. Users with interests in particular instruments may find additional data in other MET LDB collections described below.

7.1 Data Not Included in the MET Core Files

MET LDB Core Files do not include three kinds of data. (1) There are no district-level data beyond an ID for school district characteristics. (2) The MET LDB Core Files contain only limited district administrative data on schools, teachers, class sections, and students (discussed below), and at that, the data will only be for MET schools, teachers, class sections, and students within a district. The reader should note, however, that MET researchers had access to district administrative data on *all* teachers and students in MET districts for several years prior to and for two years during the MET Study (data that were used, as discussed above, to estimate value-added scores on state assessments for MET teachers). Researchers interested in these more encompassing data will find them in ICPSR Study 34798 – District-Wide Files, 2009-2014. (3) The MET LDB Core Files will not contain data from various special-purpose data collections commissioned by MET researchers and conducted with small, sub-samples of MET teachers. This includes a special study that administered the Surveys of Enacted Curriculum to a sub-sample of teachers, the Phase 1 video scores, UTOP observation scores, and a special study conducted on teachers certified by the National Board for Professional Teaching Standards. The locations of these other data are discussed below.

7.2 Overview of the Structure of the MET LDB Core Files

Figure 1 provides an overview of the MET LDB Core Files. The data structure is nested and data are easily linked across files through a systematic ID system. There is a school file (with district ID); a teacher file for each teacher; a class section file that contains data on each of the class sections taught by a teacher; a student file that contains information on students enrolled in MET teachers' class sections; a classroom observation scores file that contains segment-level data from the video-scoring; and a collection of video files for each observation session video-recorded as part of the MET Study . The structure and content of these various types of files is discussed in more detail in the sections to follow.

Figure 1: The Linked Structure of MET LDB Core Files

District/School File

This data file contains an ID for the district where MET LDB teachers taught and data on the schools where they taught. All data are from Year One of the study. The data included for schools include limited measures of school organization, student composition, and aggregated test score information. The file also includes data collected from school principals about the nature of teacher evaluation processes in a school.

Teacher File

This data file contains data on those MET LDB teachers who participated in Year One of the study only *or* who participated in both years of the study. There is one data record per teacher. Data in the teacher file was collected (or recorded) only once during the study. Among the variables included in the file are: (1) teacher IDs; (2) ID variables for district and school; (3) variables indicating a teacher's grade, subject, and study status; (4) measures of teachers' personal characteristics and professional background; (5) teacher responses to MET teacher working conditions survey [administered in Year One of the study]; (6) teacher responses to the MET teacher survey administered [in Year Two of the study]; 7) teachers' scaled scores as well as multiple choice and constructed response sub-scores for the CKT measures; and (8) principal ratings of a teacher's effectiveness.

Class Section File

This data file contains data on the focal class sections taught by MET LDB teachers. There is one data record per section. In most cases, generalists have one class section per year in the study, so that teachers who participated in both years of the study will have a total of 2 class sections of data per teacher record per year. Specialists generally have 2 class sections in Year One and 1 class section in Year Two, so that teachers who participated in both years of the study will have a total of up to 3 class sections of data per teacher record. Among the variables included for each class section taught by a teacher are: (1) section IDs; (2) ID variables for teacher, school, and district; (3) variables indicating a teacher's grade, subject, and study status; (4) measures of class composition, including aggregated data on students' prior year test scores, ethnic composition, free lunch status, and special education status; and (5) class size. Also included in this data file are: (6) value-added measures of teaching effectiveness based on student achievement scores [aggregated to the section level]; (7) measures of teaching effectiveness based on classroom observation score data [aggregated to the section level]; and (8) measures of teaching effectiveness based on student survey data [aggregated to the section level].

Student by Section File

These data files contain data on students who were in the focal class sections of MET LDB teachers during either year. Data on each student include: (1) student ID; (2) ID variables for section, teacher, school, and district; (3) measures of current and prior student achievement for all tests/years recorded [e.g., 2010-2011, 2009-2010; and up to three prior years]; (4) measures of student background [sex, ethnicity, lunch status, special education status, program participation]; and (5) student survey responses.

Segment Level Observation Scores Files

These data files contain data from all classroom observation sessions conducted on each teacher. There will be one file per observation instrument and each file will have one record for each segment scored by a rater. Data on each observation session include: (1) a segment ID, (2) ID variables for video, section, teacher, school, and district; (3) variables indicating the video's grade, and subject; (4) a variable indicating whether score comes from the primary scorer or a secondary scorer; (5) scores on the dimensions of the given observation instrument; (5) a variable indicating if the rater deferred scoring of the video to the scoring leader. Note that each video will generally have multiple rows because scores were given at the segment level. Additionally, instruments with different segment scoring lengths will not be comparable at the segment level.

Classroom Videos

These are the captured videos for each observation session in the study. These files allow researchers to see the actual classroom instruction. If a teacher fully participated in the study and if appropriate consents were obtained, each teacher would have up to four video files per section, per year. Note that some teachers did not consent to re-use of their videos by researchers who were not part of the original MET Project. Videos are available through a separate web-based video streaming system.

7.3 District/School Data File

The MET LDB Core Files include a district/school file with unique IDs for each district in the study and for each school in the study.

7.3.1 District data. The MET Core Files do not contain information about school districts other than a district ID.

7.3.2 School organization and demographics. The district/school file prepared for MET LDB Core Files contains limited data on schools. These data are only for schools that participated in the MET Study (not the broader set of all schools in MET districts), and the data are only for Year One of the study. The Core Files contain a unique school ID for each MET school that can be linked to data in ICPSR's larger MET LDB holdings.

Among the data included in the core study school file are basic measures of school student composition obtained during Year One of the study. In addition, the MET LDB Core Files list the actual grade levels at a school and the grade levels of any teachers participating in the MET Study. The school-level variables obtained from district administrative files and available in the Core Files are: (1) percentage of students of different race/ethnicities, (2) percentage of students on free lunch, special education students, and gifted status, (3) aggregated data on students' prior year test scores, and (4) school size.

7.3.3 Data on/from school principals. The MET LDB Core Files contain very limited data on school principals. No data are included on a principal's personal characteristics (e.g., age, sex, race/ethnicity) and no data are included on professional experience (e.g., degree or certification status, years of teaching experience, years of administrative experience).

7.3.4 Principal survey data. MET LDB users can find some of the data from the MET Principal Survey (administered in Year Two of the study) in the school file. The data included in the school file includes principals' responses to questions about the school's current teacher evaluation policies, about the training they received in the area of teacher evaluation, about the importance of various sources of information to their evaluations of teachers, and about their perceptions of the adequacy, accuracy, and fairness of the teacher evaluation system currently in place in the district. The school file contains the item responses of each school principal on these questions.

7.4 Teacher File

The MET LDB core study teacher file contains information on teachers that was collected once during the study. Among the variables included in the core study teacher file are social and professional background variables, variables from the teacher surveys administered as part of the study, teacher responses to and scaled scores from the CKT assessment, and principal ratings of a teacher's effectiveness.

7.4.1 ID Variables. The teacher file contains one record per teacher. Each record includes a unique teacher ID, as well as a district ID and school ID to facilitate linking to school and district level data.

7.4.2 Teacher study status. Each teacher record in the MET LDB core study teacher file includes a variable identifying the grade at which the teacher taught, a unique label that describes the "exchange group" the teacher belonged to during Year Two of the study, variables that describe the randomization block that a teacher was in (the randomization block differs from the exchange group due to attrition and the issue discussed in footnote 22), a variable indicating if the teacher was a generalist or a specialist, a variable showing participation status in the study, and information about the date at which the teacher's participation status in the study changed (if it changed)[19]. The core study teacher file also includes information about the subject(s) for which video observation data were collected on each teacher, and the number of sections from which these data came. For the convenience of users, ICPSR staff combined these data to provide users with a single indicator variable of the grade/subject combinations in which teachers participated in the MET Study (see Table 1).

7.4.3 Teacher personal/professional Background. Each teacher record in the core study teacher file also contains limited data on a teacher's personal characteristics and professional background. These data were taken from district administrative files in Year One of the study. The variables include a teacher's sex, race/ethnicity, degree status, and years of teaching experience (in total and in the district).

7.4.4. Teacher working conditions survey. As discussed earlier, MET researchers administered a MET Teacher Working Conditions survey to all teachers in MET schools during the period March through May of the first year of the MET Study (AY 2009-2010). The MET core study teacher file contains only the survey responses of teachers participating in the MET Study. Users wishing to access data from the total pool of respondents to the MET Teacher Working Conditions Survey should see ICPSR Study 34345 - Item-Level Survey Instruments and Assessment Files, 2009-2011. The Teacher Working Conditions Survey had more than 200 items asking teachers to report on many different features of their school, using many items that were borrowed from previous school surveys conducted in the United States. The core study teacher file contains the item response of teachers to this survey.

7.4.5 MET teacher survey. At the end of Year Two of the MET Study (AY 2010-2011), teachers still participating in the MET Study were administered a web-based MET Teacher Survey. The survey contained 48 items asking teachers to report about various aspects of their work with principals and of the evaluation system in place at their school. The MET core study teacher file contains the item responses of each MET teacher to each item on this survey.

7.4.6 CKT assessment. The core study teacher file contains data on CKT assessment results. The file will contain data for each CKT assessment administered as part of the study. For each subject, the data include: (1) a variable indicating the test form taken; (2) a teachers' raw score, defined as a teacher's number of correct selected-response items on the assessment, combined with a scaled

[19] In a small number of cases, three teachers were in an exchange group such that all three teachers could not be randomly assigned to a class during the same period. Randomization proceeded as described in the following example. Teachers A, B, and C are in an exchange group. There are two sections that can be assigned during period one and one section free during period two. Teachers A and B are free during both periods while teacher C must teach during period one. First, either A or B was randomly selected to teach the single section during period two. Second, teacher C and whoever was not previously selected were randomly assigned to the two sections during period one. This resulted in two separate variables describing the randomization block a teacher was in for this small group of teachers. Randomization block 1 describes the selection between A and B while randomization block 2 describes the final level of randomization.

score for constructed-response items, where constructed-response items were scaled on a 3-point range; (2) a teacher scale score, scaled to a maximum possible score of 100; and (3) a total latency score, scaled in seconds, that is equal to the sum of the selected-response and constructed-response item latencies. Users interested in item response data for the CKT assessments should see ICPSR Study 34345 - Item-Level Survey Instruments and Assessment Files, 2009-2011.

7.4.7 Principal ratings of teacher effectiveness. The core study teacher file contains the responses of a teacher's principal to a question on the MET Principal Survey that asked each principal to provide an effectiveness rating (on a 6-point scale) of up to 12 MET teachers in their school, as well as an assessment of how confident they were in this rating. Principals also reported on the number of times they observed the teacher, both formally and informally. Each teacher record in the core study teacher file includes the effectiveness rating and confidence level assigned to that teacher by his or her principal as well as the number of observations by the principal.

7.5 Section File

The MET LDB core study section file contains data on the focal class sections taught by MET teachers for both years of the study. Importantly, because of the study design, the data record for each teacher will include data on up to two class sections in Year One and one class section in Year Two[20]. The section level file has IDs to link the section to a school and teacher, variables indicating the section status at various points, as well as aggregated information on student characteristics, student surveys, student test scores, and classroom observation scores.

7.5.1 ID Variables. Given the study design, the section file IDs will be an important asset for users. The section file contains one record per section. Each record will include a unique section ID, as well as a district ID, school ID, and teacher ID to facilitate linking to school and teacher level data.

7.5.2 Section study status. Each section record in the MET LDB core study section file includes a variable identifying the grade of the section, a variable identifying the subject of the section, and a single variable indicating the grade/subject combinations in which teachers participated in the MET Study (see table 3). Also, included is a unique label that describes the randomization "exchange group" (or block) of the teacher, a variable describing the section's participation status, and the date at which the teacher's participation status in the study changed (if it changed).

7.5.3 Student Composition. Each record in the core study section file also contains limited data on the demographic composition of the students in the section. Variables include (1) percentage of students of different race/ethnicities, (2) percentage of special education students, of gifted students, and of students with free lunch status, (3) aggregated data on students' prior year test scores, and (4) class size.

7.5.4 Value Added Scores. Each section record also contains student level data aggregated to the section level (their creation is described previously). Value added test scores for the state assessments as well as the supplemental assessments are included. State value added test scores are present for both math and ELA for every section. Value added scores from supplemental tests are only available for the relevant subject (e.g. BAM for math sections and SAT9 for ELA sections). For

[20] There are seven teachers in year two who are participating with two separate sets of sections (e.g. an 8th grade section and a 9th grade section).

example, a sixth grade math classroom will have value added math scores from the state math and ELA tests and for the supplemental BAM test, but will not have scores for the SAT9 supplemental test.

7.5.5 Classroom Observation Scores. Each section record also contains aggregated data on the classroom observation instruments. Scores for the same segment by different raters were averaged to create segment level scores. The harmonic mean of these segment level scores was taken for all segments belonging to the section to obtain these section level aggregates. All sections will not have scores on all observation instruments. Only those instruments that a section could be scored on will have scores. As discussed before, UTOP scores were only collected on a small sub-sample and so are not considered part of the LDB Core Files. UTOP scores can be accessed through ICPSR Study 34346 - Item-Level Observational Scores and Supplemental Test Files, 2009-2011.

7.5.6 Student Survey Aggregates. The section level data file also contains adjusted, aggregated information on each of the previously described scales from the student perception survey. The scales included are the 7Cs (Care, Captivate, Challenge, Clarify, Confer, Consolidate, and Control), as well as scales for Reading At Home, Effort Exerted in Class, Test Preparation Practices, College Aspirations, and Happiness in class. In addition, four composite scales were created by combining the 7Cs. The first scale, called Composite1, is the average of all 7Cs. The second scale, called Composite2, is the average of all 7Cs except for Control. This scale was created a result of factor analyses showing Control did not load on the main factor. The last two composite scales are support and strictness. Support measures the degree to which the teacher helps students succeed and is composed of Care, Clarify, Captivate, Confer, and Consolidate. Strictness measures "academic press" and is composed of Challenge and Control. Data from the item responses was combined into scales at the student level then aggregated to the section level and adjusted based on classroom characteristics separately for each of the scales, as described previously.

As discussed previously, when sections taught by generalist teachers were administered the student survey, students were randomly selected to answer questions about math instruction or ELA instruction. Thus, generalist teachers have separate scores for ELA and Math. To account for this in the data file, each variable containing a scale from the student perception survey is subject specific. That is, there are three versions for each student survey aggregate in the section level file: one for Math, one for ELA, and one for Biology. Only scores for the subject of the section will be present. The distinction between the elementary version of the survey and the secondary version (discussed previously) was not retained in the classroom level aggregates, so care should be taken when doing cross-grade comparisons using the student perception surveys

7.6 Student by Section File

The MET core study student file contains person-level data on students who were in the focal class sections of any MET teacher. The student-level file contains information on students in both Year One and Year Two. Importantly, the records are *not* unique by student, but rather are unique at the *student by section* level[21]. Also, the student file does not contain data on *all* of the students a teacher taught in a given year. Rather, the student data are only for students who were enrolled in class

[21] There are a handful of students with multiple supplemental test scores on the same test for a given section. These students are not unique at the student by section level. They have one record for each time they took the supplemental test. For example, if a student took the BAM twice while in the same section, that student is listed with the same student and section ID twice. The only difference between these records is the BAM score.

sections that were a focus of the MET Study. The student file contains demographic information as well as results from the student level tests and the student survey.

7.6.1 ID Variables. The MET LDB core study student by section file contains an ID variable unique for each student as well as ID variables that indicate district, school, teacher, and section for the given student record. Generally speaking, each student will appear more than once in the student level file if a student was in more than one MET section. Records are unique at the student by section level. For example, if a student was in an Algebra I section included in the study ***and*** a biology section that was included in the study, that student is listed twice: once for the biology section and once for the Algebra I section. Moreover, it is possible for a student to be listed more than once in sections for the same subject. For example, if a student transferred sections during a year, he or she is listed in both sections. This is necessary to characterize the nesting of students within sections. The student status variables (described below) can help clarify when students were in each section.

7.6.2 Student Study Status. The student by section file contains variables indicating a student's grade, the subject of the section for the given record, if the student moved into the section after the initial rostering, and when the student was added to the section. For the convenience of users, ICPSR staff again combined the grade and subject information into a single variable (see Table 3). Further, two variables denote a student's randomization status. The first variable gives the teacher ID for the teacher to which the student was randomly assigned. This variable is blank if the student was not randomly assigned. The second variable denotes whether the student conformed to random assignment; that is, whether his or her classroom teacher was the teacher to whom that student was randomly assigned.

7.6.3 Student Demographic Data. Each student record contains district administrative data about the student. Variables include: (1) student race/ethnicity, (2) student gender, (3) special education status, (4) free lunch status, (5) gifted status, (6) English language learner status, and (7) age.

7.6.4 Student Test Scores. Each record also contains information on the test scores for students in the study. For the state tests, the scores are all presented as rank-based z-scores (described previously). All state test scores are included for each student. When a student appears more than once in the data file, the student's test scores are repeated each time the student appears. For the supplemental tests, section sub-scores, raw scores, and scaled scores are included (where available). Only the supplemental test scores for the relevant section are included for each record. Further, the supplemental test scores are only included for the section the student was in when he or she took that test[22]. For example, if a student was in section A from the beginning of the school year through April and then switched to section B, that student would have taken the supplemental test in section B. Thus, the record for that student in section B would contain information on his or her supplemental test score, while the record for the student in section A would contain missing data for the supplemental test score.

[22] Some students have multiple scores on a supplemental assessment for a given section. In the analyses for papers published by MET researchers, the scores received during different testing sessions were averaged to create a single score for a student in a section when this occurred with BAM or SAT9 scores. When this occurred for ACT scores, the first exam taken (based on the test date variable) was kept and others were dropped. All scores are kept in the Core Files to allow researchers to make their own decisions regarding these data quirks.

7.6.5 Student Survey Data. The student file also contains item responses on the student survey for both years. In addition to the student survey item-level responses, scale scores are provided based on the major scales of the survey. In all cases, the scale scores were simple means of the z-scored items loading on that target scale. The student by section file codebook has details on which items loaded onto various scales. Here, we simply name the included scales. Brief scale descriptions were presented previously. The scales included in the file are the 7Cs (Care, Control, Clarify, Challenge, Captivate, Confer, and Consolidate) and four composite scales of the 7Cs : the average of the 7Cs (labeled Composite1); the average of the 7Cs without Control (labeled Composite2); the average of Control and Challenge, which is called strictness; and the average of Care, Clarify, Captivate, Confer, and Consolidate, which is called Support. Additionally scales are called Reading At Home, Effort Exerted in Class, Test Preparation Practices, College Aspirations, and Happiness in class.

As noted before, there was a distinction between the elementary version of the survey and the secondary version of the survey. This distinction is present in the student by section file. The elementary version of the student perception survey and the secondary version are contained in separate variables. For the elementary version of the survey, there is a variable that denotes the subject students were asked to respond about when completing the survey. This variable will be important when aggregating scores to the section level because, in generalist classrooms, half the students answered regarding their math instruction and half answered regarding their ELA instruction. The secondary version of the survey does not contain a variable explicitly designating the subject students were asked to respond about while answering the survey because all students answered while thinking about the same subject.

7.6.6 Using the Student by Section File. The organization of the student level file is quite complex because of the intensive rostering done by the MET Study. Teachers had multiple chances to update rosters and list whether students were still in their classroom. This led to very detailed information of student movement between sections. Hence, students are often listed across multiple sections and often only spent part of the year in each of the sections in which they are listed. Analysis of the student level file needs to be done carefully to take into account this movement of students between sections. Use of these files could require extensive restructuring, for example, if an analyst wishes to aggregate student data across sections. Alternatively, an analyst might want to calculate the percentage of the school year that each student spent in a given section. Such a statistic would allow a researcher to look at within-school student movement across a given year or create a more precise VAM estimate that incorporates percentage of responsibility a teacher had for each student. MET researchers addressed this problem by including student scores in the aggregates of each section that students are listed in. Thus, if a student was listed in three different math sections, that student contributed to the section level aggregates of all three sections.

7.7 Observation Scores Files (Segment Level Files)

The MET core study observation scores files contain the dimension level scores from all videos scored as part of the MET core study. There are separate files for each instrument. However, the structure of the files is the same. Thus, we will only describe the general file structure. Importantly, the segment lengths were different for the different instruments. Thus, while each file contains information at the "segment" level, the meaning of segment is, generally, not consistent across files.

7.7.1 ID Variables. Each MET LDB core observation scores file contains one record for every scoring round engaged in by a rater. Thus, most segments will have one record in the file, but double scored segments will have two records. In addition to segment IDs that uniquely identify different segments

for an instrument, the files contain district, school, teacher, section and video IDs. A segment number variable indicates the order of the segment within the video (e.g. 1 means the first segment).

7.7.2 Observation Scores. An observation scores file contains dimension level scores assigned by each scorer. This includes one score for each dimension of the instrument. Only scores assigned by a rater are included. No calculated domain scores are present. For some instruments, raters assigned scores for the video as a whole (e.g. MQI raters score a best guess at teacher's MCK based on the lesson). These scores are repeated for each segment of the video (e.g. all four segments of a MQI video contain the same score for teacher MCK).

The observation scores are more complicated for the QST instrument. The QST instrument was scored as three Groups of Scales. That is, three different raters scored a segment on different parts of the QST instrument. For the QST file, the same structure is maintained, but this structure results in 8 out of the 12 scored dimensions blank in each record. Each segment, then, appears 3 times in the file (6 times if it was double scored) with a different 4 of the 12 dimensions scored on each occurrence.

7.7.3 Score Descriptors. Additional variables are present to provide more information on the observation scores. First, a rater ID is present. The rater ID is globally unique across instruments. In addition, the subject and the grade of the section being videoed are included in the file. For the QST instrument, there is also a variable indicating the Group of Scales for the given record.

7.7.4 Audio/Video Quality Scores. The last variables present in the observation scores file are scores given to audio and video quality by the rater. The rater gave three audio-visual scores on a three point scale from Low Quality to Medium Quality to High Quality. One score was given for the audio, one score was given for the board camera, and one score was given for the panoramic camera.

7.8 Video Files

7.8.1 Video Information File. The Video Information File (ICPSR # 3477) provides metadata and identifying information for videos submitted to the MET study. Each video contains footage of a single class period of a MET focal section by a MET teacher. A specialist teacher that fully participated in the study will generally have 8 videos: four from each year. A generalist teacher that fully participated in the study will have 16 videos: 8 from each year. Some teachers submitted additional videos.

Each video has a unique video Session ID that can be linked to the quantitative data through information available in the MET LDB Core Files or in the Video Information File. The Video Information File also provides the focal topic of the video and identifies videos that were master coded for training and rater calibration. (See also section 11.0 Observation Score Calibration and Validation File.)

7.8.2 Availability of Videos.

The complete set of videos used in the MET Project is not available to researchers. Some teachers who gave consent for their videos to be scored in the MET Project did not agree to provide access to other researchers. Approximately 11,500 video sessions are currently available, which is about two thirds of the sessions scored during the MET Project. The artifacts uploaded during the data collection process (e.g. lesson plan, reflection, classroom handouts and student work) are not currently available for use.

Videos collected for the MET project are provided through a secure website and may not be downloaded. The MET LDB video website is searchable by video Session ID, subject (Math/ELA/Biology), grade, period, and capture year. Users who wish to associate a video with a particular teacher, should find the Session ID in the Video Information File (Study # 34771).

7.8.3 MET Extension Project Videos.
The MET LDB video collection includes videos from the MET Extension Project (MET-X), which were captured during the 2011-2012 and 2012-2013 academic years. MET-X enlisted more than 350 teachers who had participated in the original MET Project. MET-X videos are identified as such in the Video Information File (Study # 34771). The MET-X videos are available from the MET LDB secure streaming service. ICPSR holds no associated metadata or quantitative data about these videos and cannot respond to user inquiries about their content or use. Please visit the MET-X website or contact support@umichsoe.zendesk.com for further information.

7.8.4 Video Format. MET Project videos were captured by a device that simultaneously recorded a panoramic view of the classroom and a view of the board. A small number of videos were captured on hand held cameras. The videos are stored in mp4 format. The panoramic cameras stream at a resolution of 1153x299 and the board cameras stream at a resolution of 478x338. Viewing the videos will require users to download a free Flash player.

8.0 Data Collections in the MET LDB

ICPSR has organized the MET LDB data into seven collections each of which has an ICPSR "Study Number":

- Study Information (ICPSR 34771)
- Core Files, 2009-2011 (ICPSR 34414)
- Base Data: Section-Level Analytical Files, 2009-2011 (ICPSR 34309)
- Base Data: Item-Level Supplemental Test Files, 2009-2011 (ICPSR 34868)
- Base Data: Item-Level Observational Scores, 2009-2011 (ICPSR 34346)
- Base Data: Item-Level Surveys and Assessment Teacher Files, 2009-2011 (ICPSR 34345)
- District-Wide Files, 2009-2014 (ICPSR 34798)

ICPSR Study # 34771 - Study Information
Contained in this release are a video information file, a randomization file, a subject ID crosswalk and a teacher demographics file.

- The **Video Information File** contains descriptive information about the videos captured for the MET project.
- The **Randomization File** includes district, school, section and student IDs, teacher IDs for the teacher a student was randomly assigned to, the actual teacher the student was recorded as having in October and May of that school year. Student and teacher IDs in the Randomization File were revised in 2018 to allow links to IDs used in MET analytical data files and to the ICPSR_GLOBAL_IDs in the District-Wide Files (Study # 34798; see section 9.3 Linking students and teachers in the District-Wide Files).
- The **Subject ID Crosswalk** contains only ID variables and is included to describe the associations between districts, schools, teachers, sections and students.
- The **Teacher Demographics** file contains demographic variables on only MET teachers.
- The **Student Global ID Crosswalk File** provides a way to link students across files in different parts of the MET collection. In the 2013 data release Student IDs were harmonized in the Core files, but students received different STUDENT_ICPSR_IDs in the district-wide files (ICPSR 34798). There is now one STUDENT_ICPSR_GLOBAL_ID for every student. This crosswalk associates all STUDENT_ICPSR_IDs with STUDENT_ICPSR GLOBAL_IDs.
- The **Teacher Global ID Crosswalk File** links students across files in different parts of the MET collection. Teachers were given different IDs in the district-wide files (ICPSR 34798) than in other MET datasets in the 2013 data release. This crosswalk associates all TEACHER_ICPSR_IDs with TEACHER_ICPSR GLOBAL_IDs.

ICPSR Study # 34414 - Core Files, 2009-2011
This is the collection of files described as the Core Files in this Guide.

ICPSR Study # 34309 – Base Data: Section-Level Analytical Files, 2009-2011
The Section-Level Analytical files are a merger of demographics, constructed and summary variables aggregated to the teacher section level for elementary and secondary teachers in both years of the MET study. Variables contained in the release include student race, age and other demographic variables, state test and supplemental test rankings, value-added variables and student perception survey composite measures.

ICPSR Study # 34868 – Base Data: Item-Level Supplemental Test Files, 2009-2011
Student achievement was measured in two ways -- through existing state assessments, designed to assess student progress on the state curriculum for accountability purposes, and supplemental

assessments, designed to assess higher-order conceptual understanding. The Supplemental Test release consists of both years' worth of data on the three supplemental assessments (SAT-9, BAM and ACT).

ICPSR Study # 34346 – Base Data: Item-Level Observational Scores, 2009-2011
Using panoramic digital video of classroom sessions taken of participating teachers and students, as well as teacher submitted commentary on their lessons (e.g., specifying the learning objective), trained raters scored video segments of recorded lessons based on particular classroom observation protocols. The Item-level Observational Scores release consists of files for both years' worth of data on the five primary observational protocols (CLASS, FFT, MQI, PLATO, and QST) and one protocol used on a sub-sample of MET teachers (UTOP).

ICPSR Study # 34345 – Base Data: Item-Level Surveys and Assessment Teacher Files, 2009-2011
The Item-Level Survey Instruments and Assessment release consists of four written response surveys, a teacher knowledge assessment and a survey of curriculum content of MET teacher lesson plans and other testing content. Surveys were given to principles, to gauge their knowledge of their teachers' effectiveness, teachers, to gauge their perception of their principles' effectiveness and of their broader working environment, and students, to analyze the value of student feedback on the effort to improve both teaching and learning. A teacher knowledge assessment was also conducted to test the utility of both newly developed and well established measures of teacher knowledge to predict measures of teacher effectiveness. The survey of enacted curriculum analyzes the subject content of MET teacher lesson plans, the state curriculum standards for each MET district, the state test content for each MET district and the content in the three MET supplemental tests.

ICPSR Study # 34798 – District-Wide Files, 2009-2014
The district wide files are comprised of one data file per district for school years 2008-09 to 2013-14 for a total of 36 data files. Each file contains information on each student in the school district including student demographic variables, such as race, age and gender, specialty student status variables such as free lunch, English language learner, and gifted and talented program participation, and student-level test rankings for math and reading. Also included are the aggregate means of student demographic, specialty status, and test score variables for each teacher.

ICPSR Study # 37090 Observation Score Calibration and Validation File

The Observation Score Calibration and Validation file consists of scores applied to MET classroom videos on five observation instruments: Framework for Teaching (FFT), Classroom Assessment Scoring System (CLASS), Mathematical Quality of Instruction (MQI), Protocol for Language Arts Teaching Observations (PLATO), and Quality of Science Teaching (QST). This data file has all scores assigned by raters, including scores used to evaluate raters during the scoring process.

9.0 District-Wide Files and Teacher Value-Added

9.1 Overview of the District-Wide Files

Test scores for estimating the contributions of teachers to student achievement, called "value-added," are provided in ICPSR Study 34798 – District-Wide Files, 2009-2014. These files contain information on student characteristics and standardized scores on achievement tests for math and reading for all students and teachers in each school district. Scripts used by the MET Project team at the RAND Corporation to compute value-added estimates are included with the data.

In 2018 the district-wide files were revised in several important ways.

- District-wide data were added for three years after the MET Project ended (2011-12, 2012-13, and 2013-14).
- "ICPSR_GLOBAL_IDs" were added to all district-wide files to allow students and teachers to be linked across years and to other MET LDB data.
- Variable names in all district-wide data files have been harmonized.
- We now provide two SAS-language scripts (aggregate and non-aggregate) that compute value-added measures for all districts in all years instead of district- and year-specific scripts.

9.2 Data files, missing values, test years, and grade levels

The MET LDB district-wide files merge information from four separate files (*demographics, tests, links,* and *means*) constructed by the MET team at RAND for computing value added measures. ICPSR used student IDs to merge these files into a single file for each district in each year to make them easier to disseminate. Since different types of information came from different administrative sources, the merged files often referred to different populations. For example, the *demographics* files include students who were no longer in the school district at the time of the test. Consequently, many rows in the district-wide files are filled with missing values. Many students with test scores have not been linked to teachers in the merged district-wide files. ICPSR is also aware that some students have been linked to more than one teacher for the same test. These discrepancies come from the administrative data provided by the school districts to RAND, and we are not able to correct them.

Important: Most students have two records in each district-wide file, one for ELA and one for Math. Tests for both ELA and Math are recorded on every row, but the teacher ID and other variables differ. Users should always use the variable "subject" to select which test scores to analyze on a particular record.

Users should note that the standardized tests reported in the MET district-wide data were administered at the end of the school year. Thus, a test given in 2011 was taken by students in the classroom they attended during the 2010-11 school year.

Since the value added estimation process uses the student's test score in the previous year as a control variable, every district-wide file contains test scores from at least two school years.

Variable names include the last two digits of the test year. For example, district-wide files for the 2009-2010 school year include test scores MATH_ZSCORE10, and ELA_ZSCORE10 for the current school year and MATH_ZSCORE09, ELA_ZSCORE09 for the preceding year. Some district-wide files include more than one previous year of test scores. Some district-wide files include test scores from two or three earlier years.

District-wide files may include several variables for grade level, because some districts reported grade levels separately for each test (Math, ELA). ICPSR has assigned the variable name "grade" to the grade level variable from the *links* file. Since the *links* file is the source of the teacher ID, a missing value for "grade" usually means that the student could not be matched to a teacher.

9.3 Computing value-added estimates

The value-added measures used in MET Project studies were computed by a team at the RAND Corporation, who provided SAS-language scripts to ICPSR for the benefit of future researchers. ICPSR has modified these scripts in two ways to run on the data files provided by the MET LDB.

First, the RAND scripts were designed to merge data from four files (*demographics, tests, links*, and *means*). ICPSR combined these four files into a single file to simplify the distribution of data. Code was added to the beginning of each script to split the ICPSR file into four temporary files, so that the code supplied by RAND would not need to be changed.

Second, ICPSR now provides two value-added scripts that are designed to work on all districts in all school years. These scripts differ on the inclusion of aggregate measures of classroom characteristics variables in the estimation equations and are called 'aggregated' and 'non-aggregated.' [23]
It is no longer necessary to provide separate scripts for each district in each year, because variable names have been harmonized. However, there are still some differences in the operation of the scripts between districts and years, because variables are not always available.[24] The current ICPSR scripts use the SAS macro language to automatically adjust for these differences. ICPSR file names include a part number (e.g. da34798-0025) that identifies the district and school year for each district-wide data file. (See Table A9.) The SAS value-added scripts use the part number in the file name to determine which variables are included in the estimation procedures.

Value-added estimates from the MET LDB will not exactly match the value-added scores appearing in the MET analytical files. In response to concerns of the school districts, ICPSR has modified the district-wide data to protect the confidentiality of individual students. All changes to the data were designed solely to mask the identities of students, and their impact on value-added estimates is very small. When the RAND scripts are used to compute value-added estimates, the

[23] Thomas J. Kane and Douglas O. Staiger (2008), Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation (Tech. Rep.) National Bureau of Economic Research Working Paper No. 14607; Bill and Melinda Gates Foundation (2010), Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project (Tech. Rep.) https://docs.gatesfoundation.org/documents/preliminary-findings-research-paper.pdf.

[24] The SD_AGE variable for students in district 71 in school year 2008-2009 is not used in the value-added scripts. This variable was not supplied by the school district, and it was not used in MET Project value-added estimates. ICPSR added ages to students in 2008-2009 by linking students to their records in later years in the district-wide data. This procedure worked well for grades 4-7, but few students who were in grade 8 in 2008-2009 appear in later school years.

correlation coefficients between estimates derived from the original data and the modified data are 0.98 or higher.

9.4 Linking students and teachers in the District-Wide Files

The district-wide files were revised in 2018 to make it possible to link students and teachers across years and to the analytical files in the MET LDB. IDs in the district-wide files released in 2014 were unique within years and linking across years was not possible. With support from the Bill and Melinda Gates Foundation ICPSR worked with the RAND Corporation to create "global" IDs for every student and teacher. The STUDENT_ICPSR_GLOBAL_ID and TEACHER_ICPSR_GLOBAL_ID identify the same person in every data file. The ICPSR_GLOBAL_IDs are harmonized with the ICPSR_IDs appearing in the Core files and other analytical datasets, so that linking from the district-wide files to other data in the MET LDB is now possible. The data files for 2008-09, 2009-10, and 2010-11 also include the original "ICPSR_IDs" for the benefit of researchers who worked with them before the new IDs were added.

9.5 MET student indicators

The 2018 release of the district-wide data includes variables MET_STUDENT_YR1 and MET_STUDENT_YR2, which replace the MET_STUDENT variable in the previous version of the data. The new variables identify which years of the MET Project (1 = 2009-10, 2 = 2010-11) the student was in a MET Project classroom.

10.0 Comparison of Phase 1 and Phase 2 Video Scoring

Video scoring was conducted in several phases. There was an initial summer pilot phase during which time a subset of 413 teachers with complete data had their 2000 videos scored with the CLASS protocol. These videos are sometimes called the "Plan B" sample. This scoring occurred prior to use of the web-based coding interface. Phase 1 occurred once this interface was established and scored the same 2000 videos using the rest of the observation protocols. Phase 2 scoring occurred later and focused on scoring both years of videos from teachers that were successfully randomized in Year Two. Phase 3 consisted of scoring videos that were only 25 minutes long on the FFT protocol. The training of raters and scoring of videos was managed by MET project partners ETS and Teachscape, except for UTOP which was managed by the National Math and Science Initiative.

Scoring Designs General	
Phase 1	**Phase 2**
For each instrument listed below, the videos to be scored are the 2000 videos chosen as the "Plan B" set already scored on CLASS in an extra-contract activity. The videos are half math and half ELA. All have complete data for the study, comprising prior-year VAM from the district; current-year VAM from the study; student perception survey data, and all teachers in the data set have a complete set of 4 (single subject area) or 8 (self-contained) videos. Assumptions relevant to all instruments: • Full online training and certification is assumed for all raters in the main MET scoring. • Raters will complete a daily calibration assessment. • Scoring begins at 00:00. • Scoring for each segment is done immediately following viewing of that segment (rater presses pause, codes the segment, then watches and codes the next segment). • Segments are coded as separate sets of scores (i.e. what occurs in one segment does not influence the codes of another segment). • All segments of a lesson that are scored are done so by the same rater, except for	For each instrument listed below, the videos to be scored are the "randomization" videos. These videos are the ones selected for random assignment for students and teachers in Year 2 of the MET study. Effectively, they are the "intent to treat" group. The Venn diagram in Figure 1 shows the various categories of these videos, in terms of membership in Year 1, Year 2, the randomization sample, and the Plan B sample. Boxes in blue are totals for the sample noted, and boxes in white a sub-samples of one of the overall samples. Assumptions relevant to all instruments: • Full online training and certification is assumed for all raters in the main MET scoring. • Raters will complete a daily calibration assessment. • Scoring begins at 00:00. • Scoring for each segment is done immediately following viewing of that segment (rater presses pause, codes the segment, then watches and codes the next segment). • Segments are coded as separate sets of scores (i.e. what occurs in one segment

CLASS (see Phase 2 CLASS design for details). • Double-scoring will be implemented as scores being assigned to the same video on the same instrument and group of scales by different raters with no knowledge of each other's score. • Validity scoring will be implemented as seeding through live scoring at a fixed rate videos that have a "correct" score as determined by the AP and ETS content leads. These videos will be blind to the rater, in that they will have no features that distinguish them from the live scoring. If an instrument is scored in a Group of Scales, the set of scales that comprise the full instrument has been subset into smaller groups. These groupings were formed in collaboration with the AP and content leads, in an effort to limit the cognitive load that individual raters must manage as well as to separate scales that may be challenging to score simultaneously.	does not influence the codes of another segment). • All segments of a lesson that are scored are done so by the same rater.

Classroom Assessment and Scoring System (CLASS)	
Phase 1	**Phase 2**
See CLASS Phase 2	Phase 2 design is the same as the Phase 1/Plan B design. a. Score the full randomization sample, minus the videos in the randomization sample that are also in Plan B. Score with 5% double-scoring and 5% validity scoring. No Plan B/Phase 1 videos will be re-scored on CLASS in Phase 2. b. CLASS-Upper Elementary (CLASS-UE) consists of videos from grades 4-6 and CLASS-Secondary (CLASS-SEC) consists of videos from grades 7-9. c. Score in 1 GoS with all 12 CLASS scales. d. Score the first two 15-minute segments of each video

	e. Two different raters will score the first and second segments of a single CLASS video. f. Record audio and video quality scores after all other scores are complete. g. This results in each rater recording 15 scores for a 15-minute portion of video viewing time; for each video with two 15-minute segments, 30 total scores will be recorded by two unique raters.

Mathematical Quality of Instruction (MQI)	
Phase 1	**Phase 2**
○ Score in 2 Groups of Scales: GoS1 consists of: ▪ Errors & Imprecision, ▪ Classroom Work Connected to Mathematics, and ▪ Explicitness & Thoroughness. Note that of CWCM and E&T, only one can apply depending on the content and grade level of the class—the other will be N/A. ○ GoS2 consists of: ▪ Student Participation in Meaning-Making and Reasoning, ▪ Richness, and ▪ Working with Students & Mathematics. ○ Score the 1000 Plan B math videos, plus 25% double-scoring. ○ Score the first four 7.5-minute segments of each video. ○ Record a holistic score on each scale after the four segment scores are recorded. ○ Record an Overall MQI score and a Lesson-Based Guess at MKT after all four segments' scoring is complete.	○ Score in 1 Groups of Scales consisting of all MQI scales. ○ Score the full randomization sample of math videos and re-score the portion of the Plan B/Phase 1 videos that are not already in the Phase 2 set. Score with 5% double-scoring and 5% validity scoring. ○ Score the first four 7.5-minute segments of each video. ○ Record a holistic score on each scale after the segment scores are recorded. ○ Record an Overall MQI score and a Lesson-Based Guess at MKT after all segments' scoring is complete. ○ Record audio and video quality scores after all other scores are complete. ○ This results in a rater recording 33 scores for a 30-minute portion of video viewing time.

○ Record audio and video quality scores after all other scores are complete. ○ This results in a rater recording 20 scores for a 30-minute portion of video viewing time on each of GoS1 and GoS2: ▪ 3 scale scores for each of 4 segments, ▪ 1 holistic score on each of 3 scales, ▪ 1 Overall MQI score, ▪ 1 Lesson-Based Guess at MKT, and ▪ 3 audio/visual quality scores	

Protocol for Language Arts Teaching Observation (PLATO)	
Phase 1	**Phase 2**
○ Score in 2 Groups of Scales: GoS1 consists of: ▪ Intellectual Challenge, ▪ Classroom Discourse, and ▪ Behavior Management; plus ▪ Representations of Content and various Content Domain and Activity Structure indicators. ○ GoS2 consists of: ▪ Modeling, ▪ Strategy Use & Instruction, and ▪ Time Management; plus ▪ Representations of Content and various EL Instructional Technique indicators. ○ Score the 1000 Plan B ELA videos, plus 25% double-scoring. ○ Score the first two 15-minute segments of each video. ○ Record audio and video quality scores after all other scores are complete. ○ This results in a rater recording 37 scores on GoS 1 for a 30-minute portion of video viewing time: ▪ 4 scale scores on each of 2 segments ▪ Representation of Content on each of 2 segments ▪ 7 Content Domain scores on each of 2 segments	○ Score in 1 Group of Scales consisting of: Intellectual Challenge, Classroom Discourse, Behavior Management, Modeling, Strategy Use & Instruction, Time Management and Representations of Content. No scores will be captured for Content Domain, EL Instructional Technique, or Activity Structure indicators. *[NOTE: if it is decided that some or all of the PLATO ALD scales will be used in Phase 2, the estimates for cost and schedule will require revision.]* ○ Score the full randomization sample of ELA videos and re-score the portion of the Plan B/Phase 1 videos that are not already in the Phase 2 set. Score with 5% double-scoring and 5% validity scoring. ○ Score the first two 15-minute segments of each video, recording scores after each 15-minute segment. ○ Record audio and video quality scores after all other scores are complete. ○ This results in a rater recording 17 scores for a 30-minute portion of video viewing time.

▪ 5 Activity Structure scores on each of 2 segments ▪ 3 audio/visual quality scores o This results in a rater recording 19 scores on GoS 2 for a 30-minute portion of video viewing time: ▪ 4 scale scores on each of 2 segments ▪ Representation of Content on each of 2 segments ▪ 3 EL Instructional Technique scores on each of 2 segments ▪ 3 audio/visual quality scores	

Framework for Teaching (FFT)	
Phase 1	**Phase 2**
o Score in 3 Groups of Scales. GoS1 consists of: ▪ Creating an Environment of Respect & Rapport and ▪ Using Questioning & Discussion Techniques. o GoS2 consists of: ▪ Establishing a Culture for Learning, ▪ Managing Classroom Procedures, and ▪ Communicating with Students. o GoS3 consists of: ▪ Managing Student Behavior, ▪ Engaging Students in Learning, and ▪ Using Assessment in Instruction. o Score the 2000 Plan B videos, plus 12.5% double-scoring o Score GoS1 in minutes 0-12 and 25-35. Score GoS2 in minutes 0-15 and 30-35. Score GoS3 in minutes 5-15 and 25-35. o Record audio and video quality scores after all other scores are complete.	o Score in 1 Group of Scales consisting of all FfT scales. o Score the full randomization sample, and re-score the portion of the Plan B/Phase 1 videos that are not already in the Phase 2 set. Score with 5% double-scoring and 5% validity scoring. o Score in minutes 0-15 and 25-35; this is the union of the segments from Phase 1. o Record audio and video quality scores after all other scores are complete.

○ This results in a rater recording 5 scores on GoS 1 for a 22-minute portion of video viewing time: ▪ 1 score on each of 2 scales ▪ 3 audio/visual quality scores ○ This results in a rater recording 6 scores on GoS2 for a 20-minute portion of video viewing time: ▪ 1 score on each of 3 scales ▪ 3 audio/visual quality scores ○ This results in a rater recording 6 scores on GoS3 for a 20-minute portion of video viewing time: ▪ 1 score on each of 3 scales ▪ 3 audio/visual quality scores	

Quality of Science Teaching (QST)	
Phase 1	**Phase 2**
No QST videos in Phase 1	Phase 2 design: a. Score in 3 Groups of Scales. GoS1 consists of: i. Sets the Context and Focuses Learning on Key Science Concepts, ii. Uses Representations, iii. Demonstrates Content Knowledge, and iv. Provides Feedback for Learning. b. GoS2 consists of: i. Promotes Students' Interest and Motivation to Learn Science, ii. Assigns Tasks to Promote Learning and Addresses the Task Demands, iii. Uses Modes of Teaching Science Concepts, and iv. Elicits Evidence of Students' Knowledge and Conceptual Understanding. c. GoS3 consists of:

	i. Initiates the Investigation, ii. Provides Guidelines for Conducting the Investigation and Gathering Data, iii. Guides Analysis and Interpretation of Data, and iv. Elicits Evidence of Students' Knowledge and Conceptual Understanding. d. Score all biology videos. Score with 10% double-scoring and 10% validity scoring. e. Score GoS1 and GoS2 on the first two 15-minute segments of each video, recording scores after each 15-minute segment. Score GoS3 on the entire laboratory videos, recording scores after each 15-minute segment. f. Record audio and video quality scores after all other scores are complete. g. This results in the GoS1 and GoS2 raters recording 11 scores for a 30-minute portion of video viewing time. For the GoS3 raters, the labs will be viewed in their entirety, and may vary somewhat in length. Assuming a 1-hour lab, GoS3 raters will record 19 scores for a 60-minute portion of video viewing time.

11.0 Observation Score Calibration and Validation File (ICPSR 37090)

The Observation Score Calibration and Validation file enables psychometric research on rater error. The MET Project may be the largest application of instruments designed to measure teacher effectiveness from classroom observations ever conducted. More than eight hundred raters were trained to score over fifteen thousand videos recorded by teachers in the MET Project. The result is a database of more than 2.4 million scored items from five observation instruments: Framework for Teaching (FFT), Classroom Assessment Scoring System (CLASS), Mathematical Quality of Instruction (MQI), Protocol for Language Arts Teaching Observations (PLATO), and Quality of Science Teaching (QST). See section 10.0 "Comparison of Phase 1 and Phase 2 Video Scoring" of this guide for more information about the scoring process.

Study # 37090 has all scores assigned by raters, including scores used to evaluate raters during the scoring process. Each row in the file is the score assigned to a segment of a video by a rater under one of the five instruments evaluated by the MET Project.

MET observation scores were assigned remotely using a web application supervised by ETS and Teachscape that displayed excerpts of videos and prompted raters for scores. Raters were trained on videos that had been "master scored" with "true" scores. At the beginning of every scoring session raters were assigned pre-scored "calibration" videos to assure that instruments were applied consistently. Even after they were approved for scoring, raters were occasionally given "validation" videos to be sure that their scores were consistent with expectations. See Kane, T. J., & Staiger, D. (2012). "Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains." MET Project website: http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf.

The Observation Score Calibration and Validation file is provided for research on questions like the consistency of scoring across raters. For example, these data show how often raters failed validation tests and needed to be re-trained on each item used in the MET Project. Users who want to combine observation scores based on videos with other types of MET data should use the observation scores found in the Core (ICPSR 34414) or Basic (ICPSR 34346) data files.

Appendix A. ID Types and Frequencies

A.1 Merging Files in the MET LDB

Since the MET Project included many kinds of data, many users will need to merge data from two or more files. This section provides an overview of the possibilities for linking different types of data files and the sample sizes that can be expected. The MET LDB includes IDs for District, School, Section, Teacher, Student, and Video. Most files include several IDs, and a person (teacher or student) may appear more than once in a file. For example, a teacher who taught three sections of math may appear three times in the Section Analytical File; while a teacher who taught both math and ELA to the same students may appear twice, once for math and once for ELA.

Core Files (ICPSR Study # 34414)

Table A1 shows the types of IDs that appear in the files included in the MET LDB Core Files.

Tables A2 to A4 show numbers of teachers who appear in pairs of files in the Core Files collection. Table A2 reports on Year One (2009-10) of the MET Project. The Teacher Analytic File describes 2741 unique teachers in Year One of the project. All of those teachers (2741) are described in the Year One Section Analytic Files and Student Analytic Files, but only 1580 appear in the observation score file for the Classroom Assessment Scoring System (CLASS). Similarly, Table A3 shows that there were 1902 unique teachers in Year Two (2010-11), who are all included in the Year Two Section Analytic and Student Analytic Files, but only 1208 are in the Year Two CLASS observation scores. Table A4 displays the number of Year One teachers who can be found in the Year Two instruments. Among the 2741 teachers present in the Year One data, 1887 have information in the Year Two Teacher Analytic Files and 1268 have data in the Year Two CLASS observation score file.

Base Data Files (ICPSR Study # 34309, 34868, 34346, 34345, 34798, 37090)

Tables A5 to A10 provide the types of IDs in the full range of MET LDB data files. These files include item-level data and instruments that do not appear in the Core Files.

Tables A11 to A13 report numbers of teachers who appear in pairs of files in Year One (Table A11) and Year Two (Table A12), as well as Year One teachers appearing in Year Two files (Table A13).

Table A1

Type of ID Present in MET Data Files in ICPSR Study # 34771—Study Information

		IDs Present in File							
Data type	**File Number**	**Video ID**	**DISTRICT ID**	**SECTION ID**	**SCHOOL ID**	**TEACHER ID**	**STUDENT ID**	**TEACHER GLOBAL ID**	**STUDENT GLOBAL ID**
Randomization File	da34771-0001.txt		X	X	X	X	X	X	X
Video Information File	da34771-0002.txt	X	X	X	X	X			
Subject ID Crosswalk File	da34771-0003.txt		X	X	X	X	X		
Teacher Demographics File	da34771-0004.txt		X	X	X	X	X		
Student Global ID Crosswalk File	da34771-0005.txt						X		X
Teacher Global ID Crosswalk File	da34771-0006.txt		X			X		X	

Table A2.
Type of ID Present in MET Data Files in ICPSR Study # 34414 - Core Files, 2009 - 2011

		IDs Present in File					
Data type	**File Number**	**Video ID**	**DISTRICT ID**	**SECTION ID**	**SCHOOL ID**	**TEACHER ID**	**STUDENT ID**
District/School File	34414-0001		X		X		
Teacher File	34414-0002		X		X	X	
Class Section File	34414-0003		X	X	X	X	
Student File	34414-0004		X	X	X	X	X
Classroom Observation Scores: CLASS File	34414-0005	X	X	X	X	X	
Classroom Observation Scores: FFT File	34414-0006	X	X	X	X	X	
Classroom Observation Scores: MQI File	34414-0007	X	X	X	X	X	
Classroom Observation Scores: PLATO File	34414-0008	X	X	X	X	X	
Classroom Observation Scores: QST Lab File	34414-0009	X	X	X	X	X	
Classroom Observation Scores: QST Non-Lab File	34414-0010	X	X	X	X	X	

Table A3.
Number of Teachers by Type of Data in Year 1 (2009-2010) in ICPSR Study # 34414 - MET Core Files

				Unique Teachers Year 1 (2009-2010)				
Type of Data	**Rows**	**N**	**Unique Non-Missing Year 1 Teachers**	**Teacher Analytic File**	**Section Analytic File**	**Student Analytic File**	**Classroom Assessment Scoring System (CLASS)**	**Framework for Teaching (FFT)**
Teacher Analytic File	2784	2784 Teachers	2741		2741	2741	1580	1555
Section Analytic File	6406	6406 Sections	2741	2741		2741	1580	1555
Student Analytic File	159837	137552 Students	2741	2741	2741		1580	1555
CLASS	30983	14179 Videos	1580	1580	1580	1580		1555
FFT	14427	13737 Videos	1555	1555	1555	1555	1555	
MQI	26664	6346 Videos	934	934	934	934	934	929
PLATO	14272	6789 Videos	944	944	944	944	944	944
QST-Lab	522	247 Videos	117	117	117	117	0	0
QST-No Lab	3698	924 Videos	216	216	216	215	0	0
Video Info	21983	21983 Videos	2638	2638	2638	2638	1580	1555

Table A3 continued.
Number of Teachers by Type of Data in Year 1 (2009-2010) in ICPSR Study # 34414 - MET Core Files

				Unique Teachers Year 1 (2009-2010)				
Type of Data	**Rows**	**N**	**Unique Non-Missing Year 1 Teachers**	**Mathematical Quality of Instruction (MQI)**	**Protocol for Language Arts Teaching Observations (PLATO)**	**Quality Science Teaching (QST) Lab**	**Quality Science Teaching (QST)-No Lab**	**Video Info**
Teacher Analytic File	2784	2784 Teachers	2741	934	944	117	216	2638
Section Analytic File	6406	6406 Sections	2741	934	944	117	216	2638
Student Analytic File	159837	137552 Students	2741	934	944	117	215	2638
CLASS	30983	14179 Videos	1580	934	944	0	0	1580
FFT	14427	13737 Videos	1555	929	944	0	0	1555
MQI	26664	6346 Videos	934		420	0	0	947
PLATO	14272	6789 Videos	944	420		0	0	950
QST-Lab	522	247 Videos	117	0	0		149	167
QST-No Lab	3698	924 Videos	216	0	0	149		230
Video Info	21983	21983 Videos	2638	947	950	167	230	

Table A4.
Number of Teachers by Type of Data in Year 2 (2010-2011) in ICPSR Study # 34414 - MET Core Files

			Unique Teachers Year 2 (2010-2011)					
Type of Data	**Rows**	**N**	**Unique Non-Missing Year 2 Teachers**	**Teacher Analytic File**	**Section Analytic File**	**Student Analytic File**	**Classroom Assessment Scoring System (CLASS)**	**Framework for Teaching (FFT)**
Teacher Analytic File	2784	2784 Teachers	1902		1902	1902	1280	1280
Section Analytic File	6406	6406 Sections	1902	1902		1902	1280	1280
Student Analytic File	159837	137552 Students	1902	1902	1902		1280	1280
CLASS	30983	14179 Videos	1280	1280	1280	1280		1280
FFT	14427	13737 Videos	1280	1280	1280	1280	1280	
MQI	26664	6346 Videos	770	770	770	770	770	770
PLATO	14272	6789 Videos	820	820	820	820	820	820
QST-Lab	522	247 Videos	111	111	111	111	0	0
QST-No Lab	3698	924 Videos	161	161	161	161	0	0
Video Info	21983	21983 Videos	1917	1917	1917	1917	1371	1359

Table A4 continued.
Number of Teachers by Type of Data in Year 2 (2010-2011) MET Core Files

			Unique Teachers Year 2 (2010-2011)					
	Rows	**N**	**Unique Non-Missing Year 2 Teachers**	**Mathematical Quality of Instruction (MQI)**	**Protocol for Language Arts Teaching Observations (PLATO)**	**Quality Science Teaching (QST) Lab**	**Quality Science Teaching (QST)-No Lab**	**Video Info**
Teacher Analytic File	2784	2784 Teachers	1902	770	820	111	161	1917
Section Analytic File	6406	6406 Sections	1902	770	820	111	161	1917
Student Analytic File	159837	137552 Students	1902	770	820	111	161	1917
CLASS	30983	14179 Videos	1280	770	820	0	0	1371
FFT	14427	13737 Videos	1280	770	820	0	0	1359
MQI	26664	6346 Videos	770		336	0	0	827
PLATO	14272	6789 Videos	820	336		0	0	834
QST-Lab	522	247 Videos	111	0	0		131	133
QST-No Lab	3698	924 Videos	161	0	0	131		164
Video Info	21983	21983 Videos	1917	827	834	133	164	

Table A5.
Number of Year 1 (2009-2010) Teachers Also in Year 2 (2010-2011) by Type of Data in in ICPSR Study # 34414 - MET Core Files

				Unique Year 1 Teachers Also in Year 2 Files				
	Rows	N	Unique Non-Missing Teachers	Teacher Analytic File Year 2	Section Analytic File Year 2	Student Analytic File Year 2	Classroom Assessment Scoring System (CLASS) - Year 2	Framework for Teaching (FFT) - Year 2
Year 1 Teachers	2784	2784 Teachers	2741	1887	1887	1887	1268	1268

				Unique Year 1 Teachers Also in Year 2				
	Rows	N	Unique Non-Missing Teachers	Mathematical Quality of Instruction (MQI) - Year 2	Protocol for Language Arts Teaching Observations (PLATO) - Year 2	Quality Science Teaching (QST) Lab - Year 2	Quality Science Teaching (QST)-No Lab - Year 2	Video Info Year 2
Year 1 Teachers	2784	2784 Teachers	2741	764	814	109	159	1902

Table A6.
Type of ID Present in MET Data Files in ICPSR Study # 34309 - Base Data: Section-Level Analytic Files, 2009-2011

		IDs Present in File					
Data type	**File Number**	**Video ID**	**DISTRICT ID**	**SECTION ID**	**SCHOOL ID**	**TEACHER ID**	**STUDENT ID**
Year 1 Section Level Analytical File 4th-8th Grade	34309-0001		X	X	X	X	
Year 1 Section Level Analytical File 9th Grade	34309-0002		X	X	X	X	
Year 2 Section Level Analytical File 4th-8th Grade	34309-0003		X	X	X	X	
Year 2 Section Level Analytical File 9th Grade	34309-0004		X	X	X	X	

Table A7

Type of ID Present in MET Data Files in ICPSR Study # 34868 - Base Data: Item-Level Supplemental Test Files, 2009-2011

		IDs Present in File					
Data type	**File Number**	**Video ID**	**DISTRICT ID**	**SECTION ID**	**SCHOOL ID**	**TEACHER ID**	**STUDENT ID**
ACT Quality Core series for Algebra I. English 9, and Biology - Year 1	34868-0001		X	X	X	X	X
ACT Quality Core series for Algebra I. English 9, and Biology - Year 2	34868-0002		X	X	X	X	X
Balanced Assessment in Mathematics (BAM) - Year 1	34868-0003		X	X	X	X	X
Balanced Assessment in Mathematics (BAM) - Year 2	34868-0004		X	X	X	X	X
Stanford 9 Open-Ended Reading Assessment (SAT 9) - Year 1	34868-0005		X	X	X	X	X
Stanford 9 Open-Ended Reading Assessment (SAT 9) - Year 2	34868-0006		X	X	X	X	X

Table A8.
Type of ID Present in MET Data Files in ICPSR Study # 34346 - Base Data: Item-Level Score and Test Files, 2009-2011

		IDs Present in File					
Data type	**File Number**	**Video ID**	**DISTRICT ID**	**SECTION ID**	**SCHOOL ID**	**TEACHER ID**	**STUDENT ID**
Classroom Assessment Scoring System (CLASS) - Year 1	34346-0001	X	X	X		X	
Classroom Assessment Scoring System (CLASS) - Year 2	34346-0002	X	X	X		X	
Framework for Teaching (FFT) - Year 1	34346-0003	X	X	X		X	
Framework for Teaching (FFT) - Year 2	34346-0004	X	X	X		X	
Mathematical Quality of Instruction (MQI) - Year 1	34346-0005	X	X	X		X	
Mathematical Quality of Instruction (MQI) - Year 2	34346-0006	X	X	X		X	
Protocol for Language Arts Teaching Observations (PLATO) - Year 1	34346-0007	X	X	X		X	
Protocol for Language Arts Teaching Observations (PLATO) - Year 2	34346-0008	X	X	X		X	
Quality Science Teaching (QST) - Year 1	34346-0009	X	X			X	
Quality Science Teaching (QST) - Year 2	34346-0010	X	X			X	
UTeach Observational Protocol (UTOP)	34346-0011	X		X		X	

Table A9

Type of ID Present in MET Data Files in ICPSR Study # 34345 - Base Data: Item-Level Surveys and Assessment Teacher Files, 2009-2011

		IDs Present in File					
Data type	**File Number**	**Video ID**	**DISTRICT ID**	**SECTION ID**	**SCHOOL ID**	**TEACHER ID**	**STUDENT ID**
Principal Survey	34345-0001		X		X	X	
Student Perceptions Survey - Year 1 Elementary	34345-0002		X	X	X	X	X
Student Perceptions Survey - Year 2 Elementary	34345-0003		X	X	X	X	X
Student Perceptions Survey - Year 1 Secondary	34345-0004		X	X	X	X	X
Student Perceptions Survey - Year 2 Secondary	34345-0005		X	X	X	X	X
Teacher Web Survey	34345-0006		X		X	X	
Teacher Working Conditions Survey	34345-0007		X			X	
Teacher Knowledge Assessment	34345-0008		X			X	
Survey of Enacted Curriculum	34345-0009		X	X		X	

Table A10

Type of ID Present in MET Data Files in ICPSR Study # 34798 - District-Wide Files , 2009-2014

		IDs Present in File							
Data type	File Number	Video ID	DISTRICT ID	SECTION ID	SCHOOL ID	TEACHER ID	STUDENT ID	TEACHER GLOBAL ID	STUDENT GLOBAL ID
2008-09 District 56	34798-0001				X	X	X	X	X
2008-09 District 75	34798-0002				X	X	X	X	X
2008-09 District 21	34798-0003				X	X	X	X	X
2008-09 District 18	34798-0004				X	X	X	X	X
2008-09 District 33	34798-0005				X	X	X	X	X
2008-09 District 71	34798-0006				X	X	X	X	X
2009-10 District 56	34798-0007				X	X	X	X	X
2009-10 District 75	34798-0008				X	X	X	X	X
2009-10 District 21	34798-0009				X	X	X	X	X
2009-10 District 18	34798-0010				X	X	X	X	X
2009-10 District 33	34798-0011				X	X	X	X	X
2009-10 District 71	34798-0012				X	X	X	X	X
2010-11 District 56	34798-0013				X	X	X	X	X
2010-11 District 75	34798-0014				X	X	X	X	X
2010-11 District 21	34798-0015				X	X	X	X	X
2010-11 District 18	34798-0016				X	X	X	X	X
2010-11 District 33	34798-0017				X	X	X	X	X
2010-11 District 71	34798-0018				X	X	X	X	X
2011-12 District 56	34798-0019				X			X	X
2011-12 District 75	34798-0020				X			X	X
2011-12 District 21	34798-0021				X			X	X
2011-12 District 18	34798-0022				X			X	X
2011-12 District 33	34798-0023				X			X	X
2011-12 District 71	34798-0024				X			X	X

Table A10 (continued)

Type of ID Present in MET Data Files in ICPSR Study # 34798 - District-Wide Files

		IDs Present in File							
Data type	File Number	Video ID	DISTRICT ID	SECTION ID	SCHOOL ID	TEACHER ID	STUDENT ID	TEACHER GLOBAL ID	STUDENT GLOBAL ID
2012-13 District 56	34798-0025				X			X	X
2012-13 District 75	34798-0026				X			X	X
2012-13 District 21	34798-0027				X			X	X
2012-13 District 18	34798-0028				X			X	X
2012-13 District 33	34798-0029				X			X	X
2012-13 District 71	34798-0030				X			X	X
2013-14 District 56	34798-0031				X			X	X
2013-14 District 75	34798-0032				X			X	X
2013-14 District 21	34798-0033				X			X	X
2013-14 District 18	34798-0034				X			X	X
2013-14 District 33	34798-0035				X			X	X
2013-14 District 71	34798-0036				X			X	X

Note: Use the Global IDs for linking students and teachers across files.

Table A11

Type of ID Present in MET Data Files in ICPSR Study # 37090 Observation Score Calibration and Validation File

		IDs Present in File					
Data type	**File Number**	**Video ID**	**DISTRICT ID**	**SECTION ID**	**SCHOOL ID**	**TEACHER ID**	**STUDENT ID**
Scores assigned to video segments on five observation instruments	37090-0001	X					

Table A12.
Number of Teachers by Type of Data in Year 1 (2009-2010)

Type of data	Rows	N	Non-Missing Unique Teachers	Unique Teachers Year 1 (2009-2010): Year 1 Section Level Analytical File 4th-8th	Unique Teachers Year 1 (2009-2010): Year 1 Section Level Analytical File 9th	Unique Teachers Year 1 (2009-2010): Student Perceptions Survey - Elementary
Year 1 Section Level Analytical File 4th-8th	3213	3213 Sections	2026		1	828
Year 1 Section Level Analytical File 9th	1284	1284 Sections	716	1		2
Student Perceptions Survey Elementary	19178	19178 Students	831	828	2	
Student Perceptions Survey Secondary	57200	49505 Students	1773	1094	679	3
Teacher Working Conditions Survey	15510	2072 Teachers	2072	1515	498	664
Classroom Assessment Scoring System (CLASS)	16470	7962 Videos	1580	1336	245	585
Framework for Teaching (FFT)	16470	7962 Videos	1580	1336	245	585
Mathematical Quality of Instruction (MQI)	16470	7962 Videos	1580	1336	245	585
Protocol for Language Arts Teaching Observations (PLATO)	16470	7962 Videos	1580	1336	245	585
Quality Science Teaching (QST)	641	640 Videos	231	0	230	2
ACT QualityCore series for Algebra I, English 9, and Biology	35525	28398 Students	716	1	716	2
Balanced Assessment in Mathematics (BAM)	44266	43598 Students	1288	1288	0	694
Stanford 9 Open-Ended Reading Assessment (SAT 9)	47943	47061 Students	1396	1396	1	729

Table A12 continued.
Number of Teachers by Type of Data in Year 1 (2009-2010) (continued)

		Unique Teachers Year 1 (2009-2010)				
Type of data	**Non-Missing Unique Teachers**	**Student Perceptions Survey - Secondary**	**Teacher Working Conditions Survey**	**Classroom Assessment Scoring System (CLASS)**	**Framework for Teaching (FFT)**	**Mathematical Quality of Instruction (MQI)**
Year 1 Section Level Analytical File 4th-8th	2026	1094	1515	1336	1336	1336
Year 1 Section Level Analytical File 9th	716	679	498	245	245	245
Student Perceptions Survey - Elementary	831	3	664	585	585	585
Student Perceptions Survey - Secondary	1773		1309	956	956	956
Teacher Working Conditions Survey	2072	1309		1245	1245	1245
Classroom Assessment Scoring System (CLASS)	1580	956	1245		1580	1580
Framework for Teaching (FFT)	1580	956	1245	1580		1580
Mathematical Quality of Instruction (MQI)	1580	956	1245	1580	1580	
Protocol for Language Arts Teaching Observations (PLATO)	1580	956	1245	1580	1580	1580
Quality Science Teaching (QST)	231	218	174	0	0	0
ACT QualityCore series for Algebra I, English 9, and Biology	716	679	498	245	245	245
Balanced Assessment in Mathematics (BAM)	1288	521	958	874	874	874
Stanford 9 Open-Ended Reading Assessment (SAT 9)	1396	589	1033	918	918	918

Table A12 continued.
Number of Teachers by Type of Data in Year 1 (2009-2010) (continued)

		Unique Teachers Year 1 (2009-2010)				
Type of data	**Non-Missing Unique Teachers**	**Protocol for Language Arts Teaching Observations (PLATO)**	**Quality Science Teaching (QST)**	**ACT QualityCore series for Algebra I, English 9, and Biology**	**Balanced Assessment in Mathematics (BAM)**	**Stanford 9 Open-Ended Reading Assessment (SAT 9)**
Year 1 Section Level Analytical File 4th-8th	2026	1336	0	1	1288	1396
Year 1 Section Level Analytical File 9th	716	245	230	716	0	1
Student Perceptions Survey - Elementary	831	585	2	2	694	729
Student Perceptions Survey - Secondary	1773	956	218	679	521	589
Teacher Working Conditions Survey	2072	1245	174	498	958	1033
Classroom Assessment Scoring System (CLASS)	1580	1580	0	245	874	918
Framework for Teaching (FFT)	1580	1580	0	245	874	918
Mathematical Quality of Instruction (MQI)	1580	1580	0	245	874	918
Protocol for Language Arts Teaching Observations (PLATO)	1580		0	245	874	918
Quality Science Teaching (QST)	231	0		230	0	0
ACT QualityCore series for Algebra I, English 9, and Biology	716	245	230		0	1
Balanced Assessment in Mathematics (BAM)	1288	874	0	0		658
Stanford 9 Open-Ended Reading Assessment (SAT 9)	1396	918	0	1	658	

Table A13.
Number of Teachers by Type of Data in Year 2 (2010-2011)

			Unique Teachers Year 2 (2010-2011)		
Type of Data	**Rows**	**N**	**Non-Missing Unique Teachers**	**Year 2 Section Level Analytical File 4th-8th**	**Year 2 Section Level Analytical File 9th**
Year 2 Section Level Analytical File 4th-8th	1429	1429 Sections	1423		1
Year 2 Section Level Analytical File 9th	480	480 Sections	480	1	
Principal Survey	290	290 Schools	2054	1123	338
Student Perceptions Survey - Elementary	11625	11625 Students	564	563	0
Student Perceptions Survey - Secondary	25221	23185 Students	1256	813	443
Teacher Web Survey	1826	1826 Teachers	1826	1386	421
Teacher Knowledge Assessment	1718	1718 Teachers	1718	1398	312
Classroom Assessment Scoring System (CLASS)	12594	6297 Videos	1280	1041	226
Framework for Teaching (FFT)	6294	6294 Videos	1280	1041	226
Mathematical Quality of Instruction (MQI)	2991	2991 Videos	770	651	108
Protocol for Language Arts Teaching Observations (PLATO)	3209	3209 Videos	820	693	118
Quality Science Teaching (QST)	563	563 Videos	163	0	161
ACT QualityCore series for Algebra I, English 9, and Biology	12728	11469 Students	480	1	480
Balanced Assessment in Mathematics (BAM)	21920	21481 Students	865	865	0
Stanford 9 Open-Ended Reading Assessment (SAT 9)	23175	22635 Students	923	923	1

Table A13 continued.
Number of Teachers by Type of Data in Year 2 (2010-2011) (continued)

		Unique Teachers Year 2 (2010-2011)			
Type of Data	**Non-Missing Unique Teachers**	**Student Perceptions Survey - Elementary**	**Student Perceptions Survey - Secondary**	**Teacher Web Survey**	**Teacher Knowledge Assessment**
Year 2 Section Level Analytical File 4th-8th	1423	563	813	1386	1398
Year 2 Section Level Analytical File 9th	480	0	443	421	312
Principal Survey	2054	472	941	1473	1337
Student Perceptions Survey - Elementary	564		0	526	562
Student Perceptions Survey - Secondary	1256	0		1199	1101
Teacher Web Survey	1826	526	1199		1637
Teacher Knowledge Assessment	1718	562	1101	1637	
Classroom Assessment Scoring System (CLASS)	1280	452	771	1221	1255
Framework for Teaching (FFT)	1280	452	771	1221	1255
Mathematical Quality of Instruction (MQI)	770	354	379	731	757
Protocol for Language Arts Teaching Observations (PLATO)	820	385	396	783	806
Quality Science Teaching (QST)	163	0	144	149	0
ACT QualityCore series for Algebra I, English 9, and Biology	480	0	443	421	312
Balanced Assessment in Mathematics (BAM)	865	438	397	835	849
Stanford 9 Open-Ended Reading Assessment (SAT 9)	923	465	423	895	907

Table A13 continued.
Number of Teachers by Type of Data in Year 2 (2010-2011) (continued)

		Unique Teachers Year 2 (2010-2011)			
Type of Data	**Non-Missing Unique Teachers**	**Framework for Teaching (FFT)**	**Mathematical Quality of Instruction (MQI)**	**Protocol for Language Arts Teaching Observations (PLATO)**	**Quality Science Teaching (QST)**
Year 2 Section Level Analytical File 4th-8th	1423	1041	651	693	0
Year 2 Section Level Analytical File 9th	480	226	108	118	161
Principal Survey	2054	991	597	647	115
Student Perceptions Survey - Elementary	564	452	354	385	0
Student Perceptions Survey - Secondary	1256	771	379	396	144
Teacher Web Survey	1826	1221	731	783	149
Teacher Knowledge Assessment	1718	1255	757	806	0
Classroom Assessment Scoring System (CLASS)	1280	1280	770	820	0
Framework for Teaching (FFT)	1280		770	820	0
Mathematical Quality of Instruction (MQI)	770	770		310	0
Protocol for Language Arts Teaching Observations (PLATO)	820	820	310		0
Quality Science Teaching (QST)	163	0	0	0	
ACT QualityCore series for Algebra I, English 9, and Biology	480	226	108	118	161
Balanced Assessment in Mathematics (BAM)	865	652	651	304	0
Stanford 9 Open-Ended Reading Assessment (SAT 9)	923	693	303	693	0

Table A13 continued.
Number of Teachers by Type of Data in Year 2 (2010-2011) (continued)

		Unique Teachers Year 2 (2010-2011)	
Type of Data	**Non-Missing Unique Teachers**	**Balanced Assessment in Mathematics (BAM)**	**Stanford 9 Open-Ended Reading Assessment (SAT 9)**
Year 2 Section Level Analytical File 4th-8th	1423	865	923
Year 2 Section Level Analytical File 9th	480	0	1
Principal Survey	2054	685	732
Student Perceptions Survey - Elementary	564	438	465
Student Perceptions Survey - Secondary	1256	397	423
Teacher Web Survey	1826	835	895
Teacher Knowledge Assessment	1718	849	907
Classroom Assessment Scoring System (CLASS)	1280	652	693
Framework for Teaching (FFT)	1280	652	693
Mathematical Quality of Instruction (MQI)	770	651	303
Protocol for Language Arts Teaching Observations (PLATO)	820	304	693
Quality Science Teaching (QST)	163	0	0
ACT QualityCore series for Algebra I, English 9, and Biology	480	0	1
Balanced Assessment in Mathematics (BAM)	865		365
Stanford 9 Open-Ended Reading Assessment (SAT 9)	923	365	

Table A14.
Number of Year 1 (2009-2010) Teachers Also in Year 2 (2010-2011) by Type of Data

				Unique Year 1 Teachers in Year 2 files		
Type of Data	**Rows**	**N**	**Non-Missing Unique Year 1 Teachers**	**Year 2 Section Level Analytical File 4th-8th**	**Year 2 Section Level Analytical File 9th**	**Principal Survey**
Year 1 Section Level Analytical File 4th-8th	3213	3213 Sections	2026	1416	1	1567
Year 1 Section Level Analytical File 9th	1284	1284 Sections	716	1	472	481

		Unique Year 1 Teachers in Year 2 files				
Type of Data	**Non-Missing Unique Year 1 Teachers**	**Student Perceptions Survey - Year 2 Elementary**	**Student Perceptions Survey - Year 2 Secondary**	**Teacher Web Survey**	**Teacher Knowledge Assessment**	**Classroom Assessment Scoring System (CLASS) - Year 2**
Year 1 Section Level Analytical File 9th	2026	562	807	1392	1400	1047
Year 1 Section Level Analytical File 9th	716	0	436	421	307	222

		Unique Year 1 Teachers in Year 2 files				
Type of Data	**Non-Missing Unique Year 1 Teachers**	**Framework for Teaching (FFT) - Year 2**	**Mathematical Quality of Instruction (MQI) - Year 2**	**Protocol for Language Arts Teaching Observations (PLATO) - Year 2**	**Quality Science Teaching (QST) - Year 2**	**ACT QualityCore series for Algebra I, English 9, and Biology - Year 2**
Year 1 Section Level Analytical File 9th	2026	1047	657	700	0	1
Year 1 Section Level Analytical File 9th	716	222	107	115	161	472

		Unique Year 1 Teachers in Year 2 files			
Type of Data	**Non-Missing Unique Year 1 Teachers**	**Balanced Assessment in Mathematics (BAM) - Year 2**	**Stanford 9 Open-Ended Reading Assessment (SAT 9) - Year 2**	**Balanced Assessment in Mathematics (BAM) - Year 2**	**Stanford 9 Open-Ended Reading Assessment (SAT 9) - Year 2**
Year 1 Section Level Analytical File 9th	2026	862	919	862	919

Year 1 Section Level Analytical File 9th	716	0	1	0	1

APPENDIX B: DESCRIPTION OF RANDOMIZATION PROCESS

The MET project design called for all teachers participating in the study ("MET teachers") to be randomly assigned one class of students for the 2010–11 school year. When schools joined the study during the 2009–10 school year, principals identified groups of teachers in which all teachers met the following criteria:

1. They were teaching the same subject to students in the same grade (for example, teachers teaching math to 6th graders or English language arts to 8th graders or self-contained 4th grade classes);

2. They had the necessary certification so they could all teach common classes; and

3. They were expected to teach the same subject to students in the same grade in the 2010–11 school year.

These groups of teachers were referred to as "exchange groups," and schools needed at least one exchange group with two or more teachers who agreed to enroll in the study to participate in the MET project.[1]

The plan called for identifying one class roster of students for each teacher in an exchange group and randomly assigning these rosters to the exchange group teachers. The randomized rosters would be chosen from classes of the grade-level and subject of the exchange group. For instance, if the common grade-level and subject were 8th grade math when the teacher enrolled, then only rosters for 8th grade math would be part of the randomization. We call the set of rosters that could be randomly assigned to teachers in the exchange group the "exchangeable rosters."

This appendix explains the procedures used to identify the exchangeable rosters and randomly assign rosters to MET project teachers. It also provides summaries of the randomization process.

RANDOMIZATION PROCESS

The randomization process started in early spring 2010 with the MET project gathering information from all of the partner districts on their scheduling procedures and their methods for exchanging information about assignments between schools and the district central office data system. On the basis of these meetings, the project developed a plan in which schools would complete a spreadsheet with the schedule of courses to be taught by exchange group teachers. Schools would complete the spreadsheet as soon as the schedules became available throughout spring and summer 2010. Schedules would typically be ready before the corresponding class rosters were available. Schools would send the schedules to the MET project team by deadlines dictated by timelines established by each district. The MET project team would process the schedules and make random assignments. Again, according to district timelines, districts would send the MET project team the rosters for all the classes on the schedules. When the rosters were received and verified, the MET project team would send the district and schools the teacher assignments according to separate procedures established with

1 The eligibility requirement served as a guideline and was widely but not completely enforced.

each district.[2] The timelines for completing randomization were set by each district's timeline for completing its class assignments and often required MET project to randomize rosters to teachers within a day or two after the deadline for receiving the spreadsheet schedules.

Figure A presents an example of the spreadsheet used by schools to share scheduling information with the MET project staff. The MET project prepared a custom spreadsheet for each school with the first six rows of data filled in, including the school, district, and teacher project or MET project identification numbers, the teachers' names and district identification numbers. Information filled in by the MET project team also included the exchange group identification number, the grade-level of eligible classes for the exchange group, and a subject code for the eligible subject (e.g., four for middle school math). The spreadsheet contained one page for each exchange group and a table of contents listing all the exchange groups.

Figure A

EXAMPLE OF COMPLETED SPREADSHEET TEMPLATE WITH BLOCK AND PERIOD INFORMATION FOR MET PROJECT TEACHERS

					MET Project Teacher ID	MET Project Teacher ID	MET Project Teacher ID	MET Project Teacher ID	MET Project Teacher ID
					10XX10	10XX11	10XX12	10XX13	10XX14
Dist MET Project ID	School MET Project ID	MET Project Exch. Group	MET Project Eligible Grade	MET Project Subject ID	DISTRICT Teacher ID	DISTRICT Teacher ID	DISTRICT Teacher ID	DISTRICT Teacher ID	DISTRICT Teacher ID
1	9999	DG0999	7	4	999905	999904	999903	999902	999901
					MET Project Teacher Name	MET Project Teacher Name	MET Project Teacher Name	MET Project Teacher Name	MET Project Teacher Name
Class Period	Grade Level	Course Name / Type	Course Section No.		Jane Jones	Kate Knudson	Luke Lesser	Mary May	Nate Newcomb
2	7	Standard Plus Math 7	20212000-06/363		X	X			NA
2	7	Standard Plus Math 7	20212000-12/312		X	X			NA
4	7	Standard Plus Math 7	20212000-18/375				X	X	NA
4	7	Standard Plus Math 7	20212000-24/399				X	X	NA

2 This process was not followed in one district. That district centrally managed all scheduling and could produce a data file of assignments. For that district, schools created schedules with preliminary assignments of MET project teachers and entered them into the districtwide scheduling system. Schools also entered the student rosters into a districtwide database. From its scheduling and rostering databases, the district provided the MET project with the scheduling database and rosters. MET project staff identified exchangeable classes for MET project teachers in the database and made the random assignments.

Schools added data on the teachers' schedules for eligible classes (rows 7 to 10 in Figure A). This information included the period of the day in which the class was to occur, the grade-level of the class, course name or type, and a course section number. The school also put an "X" in the rows of the column corresponding to each teacher's name if the teacher's schedule permitted him or her to teach the class during the period listed. The cells in the columns for the teacher were left blank if the teacher's schedule did not allow the teacher to teach the class. The school put "NA" in every row in a column corresponding to a teacher's name if the teacher had left the school, would not be teaching in the school or grade-level and subject in the 2010–11 school year, or requested not to be part of the MET project in year 2. The MET project team included in the spreadsheets every teacher who was participating in the study at the time that the spreadsheet was created.

Schools received detailed written instructions on how to complete the spreadsheets. Project staff also conducted webinar training for school staff on the randomization process, including how to complete the spreadsheet and how and when random assignments would be communicated with the schools. Some schools completed the spreadsheets accurately, but many made errors that project staff had to assist schools in correcting. Some schools never completed the spreadsheet and project staff, including the district liaison (or district project coordinator), needed to call these schools, obtain the information via phone, and complete the spreadsheet.

In the example in Figure A, Jane Jones and Kate Knudsen could both teach either section in period 2, but they could not teach grade 7 math in period 4. Luke Lesser and Mary May were the opposite: They could teach grade 7 math in period 4 but not during period 2. Nate Newcomb would not be teaching grade 7 math at the school in the 2010–11 school year or had decided not to participate in the study in year 2. This situation in which not all the teachers in the exchange group were scheduled to teach during a common period occurred very frequently among participating schools. To accommodate this lack of a common period, the MET project created subgroups within the exchange group of teachers who were scheduled to teach in a common period and could exchange rosters. In the example in Figure A, there would be two subgroups of the exchange group: a period 2 group with Jane Jones and Kate Knudson and a period 4 group with Luke Lesser and Mary May. These subgroups were called "randomization blocks," and rosters were randomly assigned among teachers in the same randomization block. Each teacher could belong to only one randomization block.[3] If teachers were in two or more blocks, they were randomly assigned to a block. For instance, suppose Kate Knudson could also teach in period 4 and Luke Lesser could also teach in period 2. They both would be in two possible

3 In some very rare occasions the following situation occurred:

Class Period	Grade Level	Course Name / Type	Course Section No.	Jane Jones	Kate Knudson	Luke Lesser
2	7	Standard Plus Math 7	20212000-06/363	X	X	X
2	7	Standard Plus Math 7	20212000-12/312	X	X	X
4	7	Standard Plus Math 7	20212000-18/375		X	X

There is one section in period 4 that could be taught by either Knudson or Lesser but not Jones. All three teachers can teach in period 2. There are three teachers and three sections but one teacher is not available to teach one of the sections. In this case, the project first randomly chose between Knudson and Lesser to receive the period 4 roster (say we chose Lesser) and then randomly assigned the period 2 rosters to the other two teachers (Jones and Knudson). We treat Knudson as being in two blocks: one with Jones and one with Lesser, even though Knudson only taught one randomly assigned roster.

randomization blocks and project staff would randomly assigned Knudson to one block and Lesser to the other. If only one teacher was available to teach during a period, the project called that teacher a singleton and that teacher was not randomly assigned a roster.

Within a randomization block, teachers were randomly sorted and rosters (sections) were randomly sorted and the first teacher was matched with the first roster and so on.

RANDOMIZATION SUMMARY

The project requested scheduling information for 2,462 teachers from 865 exchange groups in 316 schools. The project created 668 randomization blocks from 619 exchange groups in 284 of the participating schools. The remaining schools' schedules did not permit randomly swapping rosters among any of MET project teachers or all its MET project teachers had left the school or the study.

From these randomization blocks, the project randomly assigned rosters to 1,591 teachers.[4, 5] (This includes 386 high school teachers and 24 teachers for whom rosters were later found to be invalid.) Seven hundred, seventy teachers were not eligible for randomization because they were not scheduled to teach the exchange group subject and grade level in 2010–11 or they decided not to participate in year 2 of the study. The remaining 281 teachers could not be randomized because they did not teach in a period with two or more teachers for exchanging rosters.

4 Two teachers in blocks with a single teacher were randomly assigned rosters and counted in the randomized sample. These teachers were included in the analysis sample but do not contribute to estimates.

5 Because of a large number of teachers without exchangeable rosters in one district, the study added 33 teachers who did not participate in year 1 to the study and included them in the random assignment of the rosters. The remaining 1,558 teachers with randomly assigned rosters all participated in year 1.

Appendix C

MET Project Video Teacher Resources

Biology Lesson Selection for Teachers

The Quality Science Teaching instrument focuses on four specific dimensions of science teaching. We are interested in capturing videos of your classroom that focus on:

- Student engagement;
- Discourse based on investigative evidence and reasoning;
- Inquiry initiation, implementation, and closure. The investigation can be lab based, focus on a Science, Technology, & Society (STS) issue, and/or analysis of existing data base.
- Monitoring of student learning.

It is unlikely that any one lesson will include all the components listed above, so we recommend that across the four lessons to be videotaped, that every effort be made to provide a variety of lesson types.

1) At least one of your lessons will be recorded with a handheld device. This lesson should focus primarily on a lab or investigation activity where students are conducting the activity in pairs or small groups.

   Lab Examples involve activities where student collect actual data (either numerical or observations/drawings). Some examples include but are not limited to: examining check or onion or plant cells, osmosis of cells, diffusion rates, enzyme labs, reaction rates, surface area vs volume lab, experimenting with factors impacting plant growth rates, cellular respiration labs, recombinant DNA labs, etc.

   Investigation Activities require students to work in small groups to complete the activity, to share ideas, and to complete some type of product. Some examples include but are not limited to: creating visual models of DNA and explaining replication, working with beads or other items to model meiosis and mitosis, simulating natural selection, predatory-prey activity, models cell organelles and explain

the purposes of each, working on computer simulations, building models to represent a biological concept, debating global or societal issues, analyzing data sets and/or comparing a variety of research sources.

The other lessons will be captured using a panoramic video. For these lessons we recommend one of each of the following types of lessons

2) One lesson focusing on introducing something new to the students. This lesson might be about the introduction of a new biological concept, an introduction to an investigation or lab activity, or initiating a discussion about a global or societal issue where students can apply their understanding of biological concepts.

3) One lesson conducting a whole class discussion focusing on the analysis of data. This discussion includes: sharing the findings from a lab or investigation, analyzing the data or research gathered by the students to identify patterns or trends in the data, helping students to generate interpretations of the data, and asking students to state their conclusions based on the analysis of the data.

4) One lesson highlighting your strategies for monitoring student learning. This lesson would include ways to check for understanding that will provide you with information about the learning for the whole class and individual students. This is not intended to be a lesson focusing on administering a written test.

We are not looking for lessons that are primarily:

- Extended lectures
- Powerpoint presentations with scripted narrations
- Extended portions of showing a video or film
- Extended periods of time where students are completing In-class worksheets
- Administering a quiz or written test
- Extended periods of time devoted to test review

Appendix D

Teacher Working Condition Constructs

New Construct	Items
Facilities and Resources—Technological Resources	frl21oequip frl21relinternet frl21instrtech frl21properson frl21appmaterial frl21comm
Facilities and Resources—Physical Environment	frl21space frl21clean frl21environ
Teacher Leadership—General	eml21process eml21decmake eml21experts eml21solve eml21trustsound eml21effleader eml21tchleader
Teacher Leadership—In the Classroom	eml49instmat eml49assess eml49techniq
Teacher Leadership—School Administration	eml49schbudget eml49newtch eml49inserve eml49studiscip eml49siplan
School Leadership—General	ldl21raiseconc eml21trustresp ldl21tchrsupp ldl21sipeffect ldl21sharedvis ldl21evalconsis ldl21fdbkimpr ldl21tchrperf ldl21recogaccom ldl21profstds ldl21usedata
School Leadership—Teacher Concerns	ldl21effortld ldl21effortsc ldl21efforttm ldl21effortmn ldl21efforttl ldl21effortfr ldl21effortcs

New Construct	Items
	ldl21effortpd ldl21effortip
Instructional Practice and Support—Assessment	ipl21statedata ipl21localdata ipl21datainform
Instructional Practice and Support—Support	ipl21maxsuccess ipl21autonomy ipl21supports ipl21plcinstr ipl21trynew