Teaching Dimensions Observation Protocol (TDOP)
User's Manual

Wisconsin Center for Education Research
University of Wisconsin-Madison
http://tdop.wceruw.org
http://ccher.wceruw.org

Matthew T. Hora
Amanda Oleson
Joseph J. Ferrare

The development of the TDOP was supported by the National Science Foundation as part of the Culture, Cognition, and Evaluation of STEM Higher Education Reform (CCHER) project, which was a Research and Evaluation in Engineering and Science Education (REESE) program award (# DRL-0814724; 2008-2012)

Award No. 0814724

CCHER
Culture, Cognition, and Evaluation of STEM Higher Education Reform

Introduction

The Teaching Dimensions Observation Protocol (TDOP) is a structured classroom observation protocol designed to capture nuances of teaching behaviors in a descriptive manner. The TDOP was originally developed as part of an empirical study of the determinants of postsecondary teaching practices. Data obtained using the TDOP protocol can be used for a variety of purposes including research on classroom practice, program evaluation, faculty development, and institutional assessments. In this User's Manual we provide a brief overview of the background of the TDOP, the component parts of the protocol itself, technical information of the protocol, guidelines for use for both the hardcopy and website versions, and guidelines for data management and interpretation.

Background of the TDOP

The TDOP protocol was developed as part of the Culture, Cognition, and Evaluation of STEM Higher Education Reform (CCHER) study, which was funded by the National Science Foundation's REESE program from 2008 to 2012. The CCHER study aimed to articulate some of the cognitive, cultural, and organizational factors that shaped instructional decision-making and classroom practice. Originally, the study relied exclusively on interview and survey data, but we quickly realized that to provide rich and detailed insights about faculty behaviors, on-site observations of classroom teaching would be necessary. This perspective is based on the view that reducing classroom teaching to a single descriptor or variable, as is often done in self-reported surveys, is an insufficient way to describe this complex and multi-dimensional practice. Instead, the research demanded a way to produce robust descriptive accounts of classroom instruction that went beyond self-reported data as obtained in surveys or interviews.

In preparation for the study a variety of classroom observation protocols, used in both K-12 and postsecondary settings, were reviewed for potential adoption. Observation instruments used in postsecondary classrooms include open-ended and unstructured protocols that are commonly used in professional development and/or peer observation situations. Other more structured protocols include the Teachers Behavior Inventory (TBI), which requires analysts to rate instructors on 124 items after the conclusion of an observed class (Murray, 1983). Another widely used instrument is the Reformed Teaching Observation Protocol (RTOP), which aims to capture the degree to which instructors are using "reformed" teaching practices or not (MacIssac & Falconer, 2002). However, since the RTOP is based on underlying scales of instructional practice (e.g., classroom culture) and a priori determinations of instructional quality, the resulting data do not provide descriptive accounts of teaching, but instead pre-judge which practices are effective and which are not.

However, none of the existing protocols served our particular research goals. Unstructured protocols, while providing rich accounts of classroom contexts and teaching behaviors, could not be used to produce comparable data across courses, departments, and institutions. Criterion-reference protocols were not suitable given our goal of describing (not evaluating) postsecondary teaching practices. In particular, given findings that research-based techniques can be implemented poorly and traditional lectures delivered in a pedagogically rich manner (Saroyan & Snell, 1997; Turpen & Finkelstein, 2009), we felt that instruments that determine instructional quality prior to actually observing an instructor's practice were not suitable for our study. Further, important features of classroom dynamics and student learning are not limited to teacher behaviors, but also include other dimensions of practice such as student-teacher interactions, the cognitive engagement of students, and more generalized pedagogical strategies. Finally, these different features of instruction may unfold differently as a class proceeds, such that a single global measure may unnecessarily collapse a range of instructional behaviors into a single score. As a result, we desired a protocol that accounted for the temporal variation in how different dimensions of teaching were observed as a class proceeded. Yet no existing protocol provided a way to capture multiple dimensions of practice along these lines in a structured manner. We then set out to develop a new observation protocol that would address limitations in existing protocols while also meeting the unique needs of the CCHER study.

Instrument Development

The development of the TDOP originated with a strong foundation in a particular theoretical perspective on activity in general and on teaching practice in particular. Using this as a foundation, the instrument was adapted from an existing protocol used to study middle school science instruction, and then it underwent a series of pilot testing, reliability testing, field-testing, and revision. Because the TDOP does not aim to measure latent variables such as instructional quality, and it is not tied to external criterion such as reform-based teaching standards, some psychometric testing (e.g., factor analyses to assess construct validity) is not appropriate. However, other forms of validity and reliability testing should be a part of any instrument development process, and we pay particularly close attention to establishing inter-rater reliability (IRR) prior to the use of the TDOP to collect data about teaching.

Theoretical background

A core theoretical proposition guiding this study is that teaching is best understood as a multi-dimensional phenomenon in which various elements interact with one another over time. Thus, the common approach of conceptualizing postsecondary teaching solely as the regularity with which certain teaching methods are used (e.g., lecturing, small group work) is rejected as it constitutes an overly reductionist approach that obscures the complexity of classroom practice. This stance is partially an artifact of the reliance on questionnaires to study postsecondary teaching, and we argue that classroom observations are a far more appropriate way to measure this complex phenomenon. However, obtaining robust accounts of faculty practice requires not only improved instruments, but also conceptual frameworks with which to frame, interpret and analyze the resulting data. We suggest that an analysis of faculty teaching should be based on the premise that practice itself is best viewed as situated in and distributed among features of particular settings. Indeed, a view of teaching as solely overt teaching behaviors is based on the principle that teaching is reducible to a de-contextualized, single behavior of an individual – the lone hero premise of organizational practice that ignores critical features of the socio-structural underpinnings of work (Spillane, 2006). Instead, faculty operate in relation to the socio-cultural and structural contexts of their departments and institutions, as they are “embedded in an organizational matrix” of influences including their discipline, profession and institution. Therefore, the broader milieu of practice should be taken into account when examining faculty teaching practices (Umbach, 2007, p.263).

Research on school leadership from a distributed perspective provides a way to conceptualize educational practice that accounts for both individual agency and the context of activity. According to this perspective, educational practice is best viewed as “distributed in the interactive web of actors, artifacts, and the situation” (Spillane, Halverson & Diamond, 2001, p.23). As such, the appropriate unit of analysis for behavior is not the individual, but rather the larger activity system that encompasses individuals, artifacts, and tasks in an integrated whole (Engestrom, 1996). Thus, in order to adequately understand how faculty plan and teach their courses, it is necessary to consider how they interact with artifacts and other people within specific contexts of activity (Lave, 1988). Importantly, artifacts are not solely material objects or tools, but also encompass entities such as social structures, policies, and classroom procedures. Faculty will draw on these resources while negotiating their organizational environments and performing tasks, such that no activity is truly autonomous (Wertsch, 1991). Additionally, over time, groups will develop habits and normative expectations for particular activities, especially in the academic disciplines where a highly refined process exists for socializing students into the professions. Building on these ideas, Halverson (2003, p. 2) developed systems-of-practice theory, which focuses on the "dynamic interplay of artifact and tasks that inform, constrain and constitute local practice." Thus, in our study in general, and in developing the TDOP, we built upon Halverson's work and applied it to study two aspects of faculty practice: how these systems inform and constrain course planning, and how they actually constitute classroom instruction.

In regards to classroom instruction, teaching can be viewed in terms of participation in local networks or systems, which necessitates an account that moves beyond solely capturing the methods that teachers use in the classroom (Barab, Barnett, Yamagata-Lynch, Squire & Keating, 2002; Halverson & Clifford, 2006). Instead, teaching itself is seen as a multi-faceted practice that encompasses the teacher, students, and features of the instructional task (Cohen & Ball, 1999). Thus, classroom instruction not only includes the use of specific *teaching methods* by faculty (e.g., lecture, small-group discussion), but also the types of *cognitive demands* that teachers place on students in their class, the degree to which students are engaged in the class (i.e., *student engagement*), and the use of *instructional technology* (e.g., clicker response systems, chalkboards).[1] Later, we added the categories of *student-teacher interactions* and *pedagogical strategies* to hone in on these dimensions of practice. Each of these categories represents the core actors (i.e., teachers and students) and artifacts that comprise instructional systems-of-practice within the classroom.

In later section of this manual where we describe each of the categories, the theoretical background of some of these categories (e.g., cognitive demand) is described in greater detail, but it is important to recognize that overall each category is meant to be viewed in interaction with the others as part of an integrative system of instructional practice. The configurations that form through the collective use of the five dimensions represent "repertoires of practice" for individual instructors, departments, disciplines, and even institutions (Gutierrez & Rogoff, 2003). In the analysis below we focus specifically on these configurations at the disciplinary level since disciplines play a considerable role in shaping faculty identities, institutional structures, and approaches to teaching and learning. These configurations can be empirically discerned through techniques from social network analysis, which are increasingly being used to study complex interactions and affiliations in educational contexts (see, e.g., Daly, 2010). Given the relational assumptions of systems-of-practice theory, we argue that social network analysis is well suited to capturing the systemic nature of instructional practice.

Adaptation of the Osthoff instrument

With this theoretical framework in mind, we reviewed existing observation protocols and did not find any that were aligned with these perspectives. Fortunately, within our own institution a researcher working on middle school science instruction had developed an observation protocol that was loosely based on the Survey of Enacted Curriculum (SEC) (Blank, Porter & Smithson, 2001), which was designed to capture specific features of the content and instructional practices of middle school science instruction. Eric Osthoff and colleagues (2009) developed an observation protocol that focused on various dimensions of teaching (e.g., teaching methods, cognitive demands, etc.) as they occurred in five-minute intervals throughout a class period (Osthoff et al., 2008).

The original protocol was changed by reducing the number of categories to include only teaching methods and cognitive demand, which would capture both faculty pedagogical behaviors and student-teacher interactions, and adding a category for instructional technology. We also included a section for open-ended notes to be taken about the specific activities taking place in the classroom. Several of the specific codes used in the original instrument suited more for a middle school classroom (e.g., reading work) were eliminated, while others relevant to a university setting were added (e.g., clicker response-systems). Additionally, a review of the STEM education literature that included empirical descriptions of teaching practices provided

[1] This conceptualization is specific to our target sample for the study reported below, which consists

additional codes to include in the instrument. After the development of the original protocol, a small pilot-study was conducted as part of the CCHER study where 12 undergraduate courses were observed. As part of this process, certain codes were eliminated due to their lack of applicability to postsecondary classrooms, and others were added that were observed during the pilot study.

Selection of categories and codes

An important part of the TDOP protocol is the use of categories and codes to measure teaching practices. The categories are designed to capture distinct types or dimensions of teaching practice, and the codes illustrate different aspects of these dimensions. The categories and codes were originally selected, and have since been revised, based on a combination of adherence to theories of instructional practice, expert input and pilot testing to ensure content validity, and field experiences which have shed light on reliability issues.

The initial identification of categories and codes began with the Osthoff instrument as a starting point, and practical concerns (as outlined above) as well as a desire to ground the protocol in particular theories of instruction (e.g., Cohen & Ball, 1999; Halverson, 2003) led to initial revisions. At this early stage (i.e., fall of 2009) we also asked a group of education researchers active in math and science education, and math and science faculty, to review a proposed list of codes for each of the categories. This group of faculty confirmed that the list of codes included in the instrument were consistent with their own understanding of their teaching practice, thus providing a measure of face and content validity for the TDOP.[2] The codes and categories were further refined through a pilot study conducted for the CCHER study in the fall of 2009. An additional round of revision to the protocol occurred after data collection for the CCHER study in the spring of 2010 and then again in the spring of 2012. These revisions centered largely on the inclusion of codes that had been observed while collecting data (e.g., a code capturing student responses to questions) and through experiences during training. The revisions made during training sessions included instances where IRR could not be sufficiently established, thereby necessitating the removal of a code. In addition, emerging theoretical concerns led to the inclusion of new codes at both stages.

In its current form, the TDOP is composed of 6 categories and 46 codes. While the categories of the TDOP are fixed, users are encouraged to select the individual codes for each category that are most appropriate for a particular study or application. Thus, while a "code bank" is described in this manual that includes all 46 codes, users are not obligated to use all of them. Because the protocol does not measure any underlying scales that depend on particular codes, the inclusion or exclusion of individual codes does not harm the quality of the protocol.

Protocol testing and psychometric considerations

As with any measurement instrument, classroom observations should be designed, tested, and utilized with careful attention to the methodological quality of the instrument. Instruments that are reliable will help to ensure that trained observers will use a given protocol consistently across different teachers and contexts. The most important aspects of ensuring the psychometric quality of the TDOP included a solid theoretical grounding in the categories and codes, extensive field testing and expert consultation to safeguard construct validity, and extensive training and testing to certify IRR.

---

[2] It is important to note that the version of the TDOP that was examined by this panel of STEM education experts was an earlier version of the protocol than the one described in this manual. The earlier version had only three categories (teaching methods, cognitive demand, and instructional technology). In current field-testing we will be asking a similar panel of experts to review the categories and codes in the latest version of the TDOP to assess content and face validity.

Consideration about an instrument's validity refers to whether or not the protocol measures what it is supposed to measure or test, with different approaches to validity including criterion validity, content validity, and construct validity. Criterion-based validity refers to how well a score predicts or estimates a measure that is external to the test, and since the TDOP does not claim to be linked to any external criterion (e.g., for teaching quality), this type of validity is not relevant. In contrast, content-based validity refers to how well a measure adequately captures or measures the domain of interest (e.g., teaching, student learning, etc.). The estimation of content validity is often done by experts who vouch for the fact that an instrument does measure the domain under consideration, and we suggest that the initial consultations with STEM education and disciplinary experts provided a measure of ensuring content validity. Further, two rounds of extensive data collection have affirmed the actual presence of the codes in real classroom behavior. Another model of validity is that of construct validity, which is based on the notion that educational and psychological constructs are latent variables, and that researchers must define and then measure these unobserved constructs in a convincing manner. Again, because the TDOP does not claim to measure any latent variables, construct validity is not appropriate to test.

Reliability refers to the consistency with which a particular instrument measures the same thing each time data are collected, regardless of who is doing the data collection. Different types of reliability estimates are used to assess the quality of research instruments, including IRR that is particularly important in the case of classroom observations. IRR is used to determine if multiple observers will use the protocol in a similar fashion while scoring the same class period. In order to test IRR for the TDOP, the researchers participated in an extensive three-day training process prior to data collection. During these sessions researchers verbalized their understanding of each code and then deliberated to reach mutual understanding. In order to test this mutual understanding and establish IRR, the analysts coded three videotaped undergraduate classes (two in chemistry and one in mathematics). The results of the IRR using Cohen's Kappa are shown in Table 1 (data represent IRR testing for spring 2012 data collection – note that this version of the TDOP did not include the student engagement category).

Table 1. *Inter-rater reliability results from TDOP training in the spring of 2012*

	Teaching Methods	Pedagogical Moves	Interactions	Cognitive Demand	Instructional Technology
Analyst 1/Analyst 2	0.850	.789	.652	0.833	0.926
Analyst 1/Analyst 3	0.905	.834	.700	0.727	0.888
Analyst 2/Analyst 3	0.889	.815	.850	0.621	0.728

The IRR fluctuates according to the dimension of practice being observed. For this reason we provide the Kappa scores in disaggregated form. Note that 'cognitive demand' has the lowest Kappa scores overall, further suggesting the highly inferential nature of assessing this dimension of practice. One possible explanation for the low reliability on the cognitive demand dimension could be that the research team did not have disciplinary training in the observed classes. In an effort to increase reliability with this dimension, additional training was conducted prior to the data collection phase, though no data are available to assess the effectiveness of this additional training.

Future testing of the TDOP includes cognitive interviews with students regarding the cognitive demand category, construct validity testing with master coders, and additional reliability and field testing trials. We intend to continually test the protocol in order to improve the psychometric qualities of the TDOP.

TDOP Components: Categories

In this section we discuss in greater detail the six categories or dimensions of practice that comprise the TDOP protocol. Each category is defined, its theoretical underpinning discussed, and issues related to its operationalization in classroom observations are considered.

Teaching methods

The *teaching methods* dimension refers to the specific pedagogical technique being used by the instructor in the class. The specific measures for this construct were derived from a review of the literature on commonly used teaching methods in introductory STEM courses, as well as a review of the preliminary list by a group of STEM instructors. While these codes provide a fine-grained description of faculty members' classroom behaviors, they remain a relatively blunt measure of instruction. That is, each of these codes represents a middle-range of specificity in regards to a particular pedagogical technique, such that many codes could be further broken down into more nuanced sub-codes. For example, working out problems or computations (WP) here simply refers to whether or not the faculty member is actively solving a numerical problem in front of the class, a measure which necessarily obscures subtleties of problem solving such as specific types of problem solving procedures. However, these details can be captured in analyst note-taking if this level of nuance is desired.

Pedagogical strategies

The *pedagogical strategies* dimension refers to more broadly defined or global pedagogical behaviors that are not tied to a particular technique or method. This category is based largely on the work of Harry Murray (1983, 1999) and researchers of K-12 teaching (e.g., Ferguson, 2010) who focus on student perceptions of instructional quality. These scholars investigate certain types of teaching behaviors that are not specific to a pedagogical technique (e.g., small group work), but instead consider more global pedagogical strategies such as clarity, organizational skills, level of academic challenge, and so on. Many of the items examined in these studies are difficult to operationalize in an observation protocol, but we have found that "illustration," "emphasis," and "organization" are a few that are possible to reliably measure. Other strategies such as reading notes are based on field observations of behaviors that we thought interesting, meaningful, and measurable.

Cognitive demand

The *cognitive demand* dimension refers to inferences regarding the potential for different forms of cognitive engagement among students. It is critical to understand that this category refers to the types of cognitive engagements that the instructor makes available to students sitting in his/her classroom. It does not measure the actual types of cognitive engagement that students are experiencing in the classroom. In our opinion, such a measure cannot be done reliably or validly through third-party observations of students. After all, the actual types of cognition being experienced by students may be so idiosyncratic and/or finely grained that it is beyond the purview of currently available measurement techniques to adequately measure student cognition. [3]

In the Osthoff et al. (2009) instrument, this category is based on interpreting the instructor's intent primarily through the nature of the content being presented (e.g., a math problem, a scientific concept, etc.) and the type of teaching method being used (e.g., lecture, working out problems). This category is based on research demonstrating that the type and degree

[3] It should be noted that research in educational neuroscience is beginning to examine student cognition in real-world settings, and that it is possible in the future that a more reliable and valid way of measuring postsecondary student cognition may become available.

of student cognitive engagement in the classroom is a key feature of learning (Blank, Porter & Smithson, 2001; Blumenfeld, Kempler, & Krajcik, 2006; Porter, 2002). Measuring cognitive engagement is inherently difficult, and strategies include inferring student engagement from observations of student-teacher interactions, or of student task performance (Fredricks, Blumenfeld & Paris, 2004). For example, Nystrand and Gamoran (1991) distinguished between substantive (i.e., sustained and deep engagement with material) and procedural (i.e., compliance with classroom rules) types of cognitive engagement, which were inferred from the type and quality of classroom discourse. The categories in the Osthoff et al. (2009) instrument, as well as those included in the TDOP, are loosely modeled on Bloom's taxonomy for educational objectives (see Krathwol, 2002).

Despite challenges associated with inferring student cognitive engagement, we felt it an important dimension of instructional practice to attempt to capture. To develop these codes, we adapted the category of "cognitive demand" from the Osthoff et al. (2009) instrument and adapted them to fit the undergraduate classroom. Because this category is what Murray (1999) calls high-inference, this category received significant attention during the instrument-training phase to ensure that all analysts coded cognitive engagement in a similar fashion. Towards this end, we developed coding rules that could be followed during data collection. For example, the "connecting to the real world" code was only applied when the instructor, using an anecdote or extended illustration, linked the course material to events, place, objects or persons associated with popular culture or the state or city where the institution was located,. Another example involves the problem solving category, which was applied in cases when instructors verbally directed students to participate in a computation or other problem solving activity, usually at their own desks or in small groups.

Student-teacher interactions

The *student-teacher interactions* category is based on the premise that the degree of student engagement with instructors and/or with one another is an important component of the learning environment. In particular, we desired to capture two facets of these interactions: (1) who is doing the speaking – the teacher or the student(s), and (2) the types of questions posed. One of the sources informing this category is the Flanders Interaction Analysis (FIA) approach, which was developed for use in K-12 classrooms but has been adopted by some postsecondary researchers and faculty developers. The FIA is based on the assumption that interactions between students and teachers represent a key aspect of effective classrooms (Amidon & Flanders, 1967). The protocol distinguishes between two types of "talk" in the classroom: (a) teacher talk which is either direct (i.e., teaching methods such as giving directions) or indirect (i.e., praising, asking questions), and (b) student talk which is considered either as a "response" (i.e., convergent answers to posed questions) or an "initiation" (i.e., divergent questions, or responses to posed questions that depart from the flow of the conversation). Analysts code each type of student and/or teacher-talk every 3-5 seconds, and the intersection between code types represent the "interaction" in the classroom. These data are entered into a matrix and analyses include the amount of time each party talks during a class and the nature of student-teacher interactions. While the FIA has mostly been used in K-12 research, Gilbert and Haley (2010) argue for its use in postsecondary settings because data obtained with the protocol are easy to interpret and thus can be useful for professional development. While the TDOP does not code teacher or student "talk" at such a fine grain size, the basic idea of capturing the temporal fluctuations of who is speaking, and the nature of the subsequent interactions, is maintained in the TDOP protocol.

In operationalizing the types of questions that are posed to students (and teachers), we drew on research on how to conceptualize instructor questions from the English as a Second Language (ESL) literature (e.g., Morell, 2004). For example, in this literature some types of questions are considered rhetorical (i.e., where no answers are expected by the interlocutor) or display (i.e., where original information is sought) questions. The different types of questions are viewed as important because they provide clues about the intentions of the speaker regarding expectations for the types of responses they desire, and they are an important feature of any social interaction. Prior versions of the TDOP attempted to capture whether questions were conceptual or algorithmic in nature, but adequate IRR could not be established for this finer level of distinction. In any case, this category provides insights into both the agent who dominates the classroom discourse (i.e., teachers or students) as well as the nature of the questions being posed.

Student engagement

An important decision that protocol designers must make is whether to focus on the behaviors of the instructor, the students, or both. In creating their observation protocol (the O-TOP), Wainwright, Flick and Morrell (2003) argue that a focus on only one of the parties will necessarily ignore one of the key partners in the teaching and learning dynamic. In many cases it is the student that remains invisible during classroom observations. As Good and Brophy (2000) noted in their discussion of effective teaching in K-12 schools:

> "…observers often try to reduce the complexity of classroom coding by focusing their attention exclusively on the teacher… but it is misplaced emphasis. The key to thorough classroom observation is student response. If students are actively engaged in worthwhile learning activities, it makes little difference whether the teacher is lecturing, using discovery techniques, or using small-group activities for independent study." (p.47)

It is important to note, however, that including both instructor and student behaviors within a single protocol adds to the demands placed on the observer, particularly if these behaviors are conceptualized and measured at a finely-grained level. Furthermore, it is difficult to ascertain from an outside analyst's perspective whether or not students are actually engaged with the class and in what fashion. This is particularly true given certain conceptualizations of student engagement such as Bloom's taxonomy, whose categories are such that determining their existence, whether from student self-reports or external observers, is quite challenging indeed. Despite these challenges, we felt it important to capture in some fashion the degree to which students are engaged and/or paying attention during the class period. We thus decided to include a coarser measurement of student engagement. This measure is simply an estimate of the percentage of students who are in the immediate vicinity of the observer that are clearly paying attention to the instructor's behaviors. In developing this procedure, we drew upon the idea of a "transect" from biological research where a given area is demarcated for in-depth analysis. In this case, the analyst identifies 8-10 students to observe who are sitting within visible range, such that the eye gaze and behaviors can be clearly identified. The criteria that we use to determine "attention" is if the student is actively writing notes about the class topic and/or looking at the instructor and/or the primary activity of the class (e.g., demonstration, video, etc.). We recognize that this is an imperfect measure of engagement, and that students can be superficially engaged in the class while taking notes or looking at the instructor, but capturing (at the very least) the physical markers of attention and focus is a beginning step towards measuring this critical dimension of teaching and learning.

Instructional technology

The sixth and final category of codes in the TDOP refers to the instructional technologies used by instructors. This group of codes was included given the increasing prominence of instructional technology in the postsecondary classroom. Specific tools were identified first by a review of the disciplinary literature in math and physics, and then through a pilot study in the fall of 2009, where the actual objects and structural features used by respondents were included in the final instrument. The instructional technology category is limited to those materials or technologies used by teachers alone, such that any student-based technology (e.g., a laptop used for taking notes) is not recorded. The only exception is clicker response systems that typically involve instructors generating and posting questions while students answer them using a handheld device.

TDOP Components: Individual Codes

The six categories outlined above are then broken down into more specific codes that represent the actual instructional behaviors that are measured by the TDOP. In this section, each code is defined and briefly described. Note that the abbreviation for each code is what actually appears on the TDOP in both hardcopy and web-based versions. Also, each code is applied within the 2-minute interval in which it is observed – in cases where they bridge two intervals the code is applied once in each interval.

__Teaching Methods__

L **Lecture**: The instructor is talking to the students and not using any visuals or demonstration equipment.

LPV **Lecture with pre-made visuals**: The instructor is talking to the students while using pre-made visual aides, such as slides, transparencies, posters, pre-written chalkboard notes, etc. The instructor must be referring to topic contained in the visual within the coded time segment.

LHV **Lecture with handwritten visuals**: The instructor is talking to the students while actively writing and presenting notes, creating charts/diagrams, etc. (must either be writing or referring to what they are writing).

LDEM **Lecturing with demonstration of topic or phenomena**: The instructor uses equipment (e.g., lab equipment, computer simulation, or other physical objects other than handwritten visuals) to convey course content. The objects must be actively referenced by the instructor. (Note: this will always be co-coded with IL and CN)

LINT **Interactive lecture:** The instructor is talking to the students while asking multiple, successive questions to which the students are responding, and student responses are either guiding or being integrated within the discussion. (2+ rounds of dialogue; a round equals at least one relevant student response to instructor)

SGW **Small group work/discussion**: Students form into at least 2 groups of 2+ for the purposes of discussion and/or to complete task.

DW **Deskwork:** Students complete work alone at their desk/chair.

CD **Whole class discussion:** Instructor initiated/prompted discussion where students are answering and asking questions amongst themselves for a sustained period of time. This is different than an interactive lecture in which the instructor is directing all of the questions. This code is also different from small group work/discussion because conversations are not in groups but involve the entire class in a single conversation.

MM **Multimedia**: The instructor plays a video or movie (e.g., Youtube or documentary) without speaking and the students watch (instructor not speaking). If the instructor is talking extensively while using multi-media, then also code LPV.

SP **Student presentation**: The students are giving presentations to the class or otherwise acting as the primary speaker or instructor in the classroom. (Only select this code and none others as long as the primary instructor is not actively teaching the class. That is, do not switch coding to what the student is doing – just use this code and no others until the primary instructor returns.)

__Pedagogical Moves__

MOV **Moves into audience:** The instructor walks up aisles or enters the student seating area of classroom.

HUM **Humor**: The instructor tells jokes or humorous anecdotes; this code requires laughter from students – at least 2 students must laugh.

RDS **Reads**: The instructor reads verbatim from prepared notes, text, or PowerPoint slides. (Must be **extensive** reading and not just reading slide headings or definitions and then elaborating extemporaneously.)

IL **Illustration**: The instructor uses real-world examples or illustrations to demonstrate, show, or otherwise convey course content. Anecdotes and stories that are not substantive demonstrations or illustrations of the course material should not be coded. Extended thought experiments, if they include the illustration of abstract content, may also be coded here.

ORG **Organization**: The instructor writes, posts, or verbally describes the outline of class and/or clearly indicates transition from one topic to the next, including transitions from previous class to the present class (end of class overview). These transitions can be between large topics or sub-topics – the main point is that students are being alerted to a shift in focus. (This can be a brief statement, so be careful to watch for these organizational markers.)

EMP **Emphasis:** The instructor clearly states relative importance – that something is important for students to learn or remember. (This includes statements about things being important for exams, future careers, and the course as a whole. This does not include “negative” emphases, such as “you don’t need to know this” types of statements.)

A **Assessment**: The students take a test or quiz. This includes the use of clickers to answer questions that explicitly seek content-related knowledge from the students. (Verbal questions are not coded under this code, but are captured in the interactions category.)

AT **Administrative task**: The instructor and/or students make announcements, discuss upcoming assignments or exams, or engage in other logistical tasks.

Student-teacher interactions

RQ **Instructor rhetorical question**: The instructor asks a question without seeking an answer and without giving students an opportunity to answer the question. (Instructor waits less than 5 seconds for an answer – if they wait longer it is a display question.)

DQ **Instructor display question:** The instructor seeks a specific factual or conceptual answer, or asks students to solve a computational problem or a conceptual dilemma.

CQ **Instructor comprehension question**: The instructor checks for understanding (e.g., “Does that make sense?”) and pauses for at least 2 seconds, thereby indicating an opportunity for students to respond.

SNQ **Student novel question**: A student poses a question to the instructor that seeks new information (i.e. not asking to clarify a concept that was previously being discussed).

SCQ **Student comprehension question**: A student poses a question to the instructor that seeks clarification of a concept that is part of the current from the past class period, or about other topics that have already been covered.

SR **Student response**: A student responds to a question posed by the instructor. (This does not include responses to instructor comprehension questions.)

PI **Peer interactions**: 2+ pairs or groups of students are speaking to/with one another about topic. This will be co-coded with SGW but not LINT, as the latter will be guided by instructors.

Cognitive Demand

RRI **Recall and retain information:** Instructors provide **verbal or handwritten** definitions of terms or equations, or students are asked to define a term or recall basic facts through a verbal question or clicker question.

PS **Problem solving:** Instructors ask students to actively solve a problem or computation. This includes computations or evaluations of conceptual dilemmas, and is evident through explicit verbal requests to solve a problem, or to engage in thought experiments or conceptual dilemmas that require students to consider alternatives and identify solutions. This may also include verbal requests to "look at the dataset and identify patterns."

CR **Creating**: Instructors ask students to engage in creating their own ideas or products, as indicated by instructors clearly stating that students should be creative and/or generate their own ideas and products. The outcome is open-ended rather than fixed.

CN **Connections to the real world**: Students make connections between the course material and their daily lives, as indicated by instructors using physical demonstrations or verbal illustrations that link material to popular culture, the local environment, etc. The connections made may be very brief, and these connections may also include similes and metaphors, as long as they are clearly intended to make the abstract or conceptual more concrete for students.

Student Engagement

VHI **Very High:** More than 75% of the students in the immediate area of the observer are either (a) actively taking notes, or (b) looking at the instructor/course materials

HI **High:** Between 50% and 75% of the students in the immediate area of the observer are either (a) actively taking notes, or (b) looking at the instructor

MED **Medium:** Between 25% and 50% of the students in the immediate area of the observer are either (a) actively taking notes, or (b) looking at the instructor

LO **Low:** Less than 25% of the students in the immediate area of the observer are either (a) actively taking notes, or (b) looking at the instructor

Technology

PO **Poster:** Posters such as the periodic table or a map of tectonic plates.

B **Book (s):** Books used during the class period.

N **Notes:** Lecture notes actively used by instructor during the class.

P **Pointer** (e.g., laser pointer, metal pointer)

CB **Chalkboard/white-board**

OP **Overhead/transparencies** (e.g., includes slides projected using a projector)

PP **Powerpoint or other digital slides**

CL **Clickers**

D **Demonstration equipment**

DT **Digital tablet:** This refers to any technology where the instructor can actively write on a document or graphic that is being projected onto a screen. This includes document cameras, as well as software on a laptop that allows for writing on pdf files. (Note: transparencies being written on are coded as OP)

M **Movie, documentary, video clips, or Youtube video**

SI **Simulation:** Simulations can be digital applets or web-based applications.

WEB **Website:** Includes reference to course website or other online resource (besides Youtube videos) as active part of instruction.

Best practices in using classroom observations

Before describing how to use the TDOP we briefly review some best practices in using classroom observations. To contextualize these practices, first we describe two of the most common uses for classroom observations (professional development and teaching assessment) and some key issues that should be considered in both settings. Then we focus on perhaps the single most important issue facing the use of observations – that of training. We cannot emphasize enough the importance of training and the testing and establishment of adequate IRR among team members if working with multiple researchers.

Types of applications: Professional development and teaching assessment

While classroom observations can be used for a variety of purposes in postsecondary settings, there are two applications that are the most common: to support professional development activities and to assess and/or evaluate teaching quality.

Professional development. Classroom observations are often used by faculty developers as part of coaching and mentoring efforts. Some teaching and learning centers offer coaching and mentoring services where a trained faculty developer will observe a class, often with a structured protocol, and then meet one-on-one with the teacher. Importantly, this type of coaching can be conducted with a variety of instructional roles, including graduate teaching assistants, tenure-track faculty, and contract lecturers (Sorcinelli & Sorcinelli, 1988). Faculty developers will often not simply conduct a single observation as part of this process, but will integrate pre- and post-class interviews or coaching sessions where targeted feedback is provided to the instructor. One of the most important aspects of using observations for faculty development is to develop a sense of mutual trust between faculty and analysts (Millis, 1992; Chism, 2007). For instance, the University of Washington Center for Instructional Development and Research (2012) emphasizes that point by stating that center consultants "can provide [instructors] with a neutral, non-threatening perspective on [their] teaching and help [them] reflect on whether, where and how to make changes." The emphasis of fostering a sense of trust is important given prior evidence that some faculty resent being observed, coached, and generally told how to teach (Eison, 1988; Millis, 1992).

In terms of the protocols used for professional development purposes, a common approach used by campus teaching and learning centers is to utilize unstructured protocols where observers take notes during the observed class, but with no pre-specified direction about what behaviors or facets of teaching to record and in what fashion. Such approaches generally yield rich contextualized information about the observed class, yet the variability and lack of standardized data collection procedures of the approach negates the possibility of comparing data across raters or even across cases rated by the same observer. In cases where faculty developers use structured protocols, those that do not require an evaluative judgment on the part of the observer may be preferred. The rationale behind this perspective is that for instructors interested in improving their teaching, "it is much more helpful to identify what [they] are doing instead of what [they] should be." (Calvin College, 2012). Additionally, the act of judging instructional quality through the use of protocols that purport to measure teaching efficacy may engender resistance by faculty, particularly if the observers are not disciplinary experts in their field (Centra, 1979).

Assessment and/or evaluation of teaching. Another use of classroom observations in postsecondary settings is the assessment and/or evaluation of teaching for the purposes of employee performance reviews. Such evaluations are commonly conducted for peers within a given academic department, and as such, are often called peer evaluations (Braskamp & Ory, 1994). One argument in favor of peer observations rather than an outside evaluator is that only colleagues are in a position to make judgments regarding the instructor's mastery of the content and the appropriateness of the pedagogy for different types of students (Cashin, 1989; Brent &

Felder, 2004). When observation data are used for performance evaluations, they are commonly integrated with other types of data as part of a multi-faceted evaluation process. As with professional development applications, the types of protocols used in evaluation settings vary from unstructured note-taking to structured protocols.

It is important to recognize that using classroom observations as a form of "high-stakes" assessment and/or evaluation presents a number of problems – a lesson that has been learned in K-12 settings. In these latter settings, classroom observations have often been used as a complement to value-added metrics. However, recent research indicates that while value-added is a limited and variable measure of teacher quality (e.g., Rothstein, 2010), classroom observations can be equally unreliable (Guarino & Tracy, 2012). Thus, some have argued that no single data source should be used to evaluate instructional quality (e.g., Shulman, 1987). However, it is not entirely clear what the "value-added" of multiple measures is if such measures are unreliable. At the very least, as higher education moves towards establishing procedures for assessing instructional quality, it is important to learn from the ways in which such procedures have had mixed results in K-12 settings and in some ways have led to undesirable consequences for teachers' professional autonomy.

Training and preparation

As noted above, training and preparing analysts prior to entering the field to collect data is essential. The only exception to this rule is if observations are being conducted by individual analysts solely with the aim of collecting in-depth notes or other classroom-related data about an instructor. This may be done in cases of ethnographic research or faculty development. For purposes of research or more systematic professional development, however, it is highly recommended to use more structured instruments, in which case training is extremely important. This is particularly the case when observation data are being used to compare instructors over time or across courses. The training procedures for the TDOP entail three critical steps: (1) becoming familiar with TDOP codes, (2) practice coding and discussion, and (3) IRR testing.

First, the research team is provided with a list of TDOP codes to carefully scrutinize and even memorize. During actual coding it is not advisable that analysts are unfamiliar with the code definitions and abbreviations. While in-class coding is aided by having a hardcopy print out of code definitions, they should be used only as a reminder and as a last resort. Thus, prior to the second step of training, all analysts should be familiar with each code and its definition. Importantly, any questions that arise during this phase should be noted and discussed among the group, such as ambiguities or confusion regarding individual codes.

Second, the research team should then conduct "practice" coding sessions using videotaped lectures. This process could entail coding 10-15 minute segments and then pausing to discuss what was coded, when, and why. Any disagreements should be resolved, and the group should come to a consensus about what codes should have been applied to the segment. During these sessions researchers should also verbalize their understanding of each code so that it is apparent when disagreements or misunderstandings exist. The research team should then do up to 3 more "practice" coding sessions followed by group discussion, making sure to select videos that are similar to the ones that will be observed for the actual data collection (e.g., course level, discipline, etc.).[4] Given the almost infinite types of instruction that can be observed in the field, it is highly likely that questions will arise regarding which TDOP code to apply to particular situations, or it will happen that no TDOP code exists to capture a particular behavior. In an attempt to minimize confusion that may arise in these eventualities, we are in the process of preparing a list of "coding rules" that should be applied to certain situations.

[4] Unfortunately, most videotaped classes do not include adequate depictions of student behaviors to code the "student engagement" category, so this will need to be done in other ways, such as role-playing among group members or live coding in local classrooms.

Third, after the group has reached consensus on when to apply the TDOP codes in particular situations, the group should formally test IRR. This can be done using either the hardcopy or the web-based versions of the TDOP, though the web-based version is highly recommended because the calculation of Cohen's kappa is automatized. For this training select at least 3 video taped classes and code each in its entirety. Then, test IRR between each pair of analysts. Cohen's kappa should be examined for each pair on each of the 6 TDOP categories, with the goal of attaining at least .75 or higher for each category. In some cases attaining such an IRR does not happen within 3 coded classes, in which case a group can continue practice coding until IRR is improved, or simply report the resulting kappa scores with the knowledge that some may view lower scores as a severe limitation to the data. In any case, make sure to do this testing and report all results in any reports, articles, or presentations.

How to use the TDOP

The TDOP is available in two versions: a hardcopy and a web-based version. While either version is feasible for use, we strongly recommend the web-based version given the capabilities of the website to customize the protocol, automate data collection and management, and to calculate IRR.

General procedures for both hardcopy and web-based coding

The TDOP is comprised of 4 main sections: observation characteristics, classroom characteristics, 2-minute interval coding, and 2-minute interval note-taking.

Observation characteristics. In order to keep track of unique observations, it is necessary to note characteristics of an observation including observer name, instructor name, course and department name, and so on.

Classroom characteristics. Certain characteristics of the classroom should be noted prior to the actual coding of the lesson. The TDOP has fields for variables such as class size and physical layout of the room. For open-ended response items (e.g., layout), take notes about interesting features of the classroom that are salient to your study and/or catch your attention as being potentially influential regarding the class (e.g., round tables designed for small group activities).

2-minute interval coding. Once all of the observation and classroom characteristics have been recorded, it is time to start collecting classroom observation data. Once the class has officially started, as indicated by the instructor beginning to speak or otherwise clearly marking the start of the class, you should start coding. The coding process simply involves marking (either with pen/pencil on the hardcopy, or by a mouse-click on the web-version) a particular code when it is observed. A general rule of thumb is that a behavior needs to occur for more than 5 seconds to be coded. If a behavior lasts past the conclusion of one interval and runs into the next, make sure to code it twice. That is, once a new interval is started, consider it a "blank slate" and re-code all observed behaviors. Also, it is useful to have a hardcopy print-out of the code bank (all codes and their definitions) nearby while you are coding, regardless of the type of TDOP you are using (hardcopy or web-based).

2-minute interval note-taking. During the coding process there will be aspects of the class that are of interest, but are not captured by the TDOP codes. These may pertain to the course material, certain interactions between students and instructors that are not reflected in any code, and so on. For these events and occurrences, take detailed notes while you are engaged in the actual coding process. Make sure to link the notes to a particular interval so that you can later identify which TDOP codes are associated with the notes.

Special procedures for the web-based version of the TDOP

The web-based version of the TDOP can be used by two types of users: administrators and researchers/users. Each study or set of observations will need an administrator to set up observation protocols, customize the TDOP if desired, and to generally handle all IRR testing and data functions. Users will not have access to the administrative functions of the website such as creating new groups and studies, deleting observations, and so on.

Administrator functions. To sign up as an administrator go to the TDOP website (http://tdop.wceruw.org) and scroll to the bottom of the page. Click on the link for administrator sign ups. This will direct you to a page titled "Create a new administrator account" where you fill out the form, selecting a user name and password. Once on the website, you should go immediately to the "account management" page where you will see links for group and study management. The difference between groups and studies is important for the use of this website. Basically, groups are collections of researchers that can be assigned to multiple studies, whereas studies are collections of observations that are linked to a particular group. Thus, a single group can collect data for multiple studies (e.g., CCHER study, TPDM study, etc.), but the data for each study will be managed separately. From the observation end, each observation will be linked to a specific study name and ID number, and each study will be linked to a group of researchers. Thus, each group (e.g., Joe, Amanda, and Matt) can have multiple studies (i.e., sets of observations) linked to them. Calculating IRR and downloading data is done by accessing the group page, where each of the studies linked to that group (e.g., CCHER study, TPDM study, etc.) can be accessed.

Groups. Groups are collections of users that can be assigned to one or more studies. A group is given an ID number and name that you provide. Once on the "Account management" page you can click on "group management" to create, edit, or manage groups. At the "Current group" page you can add new groups, which requires assigning a group name and users. To add users to the group click on the "add user" button. If the user is already in the system, you can search for their username or create a new user for the group.

Studies. Studies are collections of observations. Details about creating and managing studies are described below.

Basic functions. Each page of the website has the following tabs at the top: "my observations," "user guide," "account management," and "log off." Each of these tabs will direct you to the key areas of the website. Some of the specific functions you will want to utilize are described in greater detail below.

Creating new studies and customizing the TDOP. From the "Account management" page you can click on "study management" to create, edit, or manage studies. At the "Studies" page you can add new studies, which requires assigning a study name and assigning the study to a group of users (see above). You will want to create new "studies" for each set of observations conducted for a particular research project, evaluation, or assessment. Also, if for some reason you wish to delete or remove a study, click on the link next to the study name.

When you create a new study you have the option of customizing the TDOP to include or exclude particular codes. If you want to add new codes, click on the "add codes" link at the top of the page and add a new code in abbreviated form in the box – also select the code category to which the code belongs. If you want to remove codes, scroll down and select codes that you wish to delete, and click "remove" at the bottom of the page. You can also edit or add codes later by clicking on the "edit instruments" link next to each study that is listed on the main study page.

Creating a new observation. Navigate to the "my observations" page by clicking on the tab at the top of the website. This is the page where you will create new observations, manage your observations, and initiate the data collection process (i.e., coding an actual class). First you will click on the "ADD" button to the right of the page to create a new observation – you will need one for each class that you observe. Here, fill in all fields that are relevant to your study, making sure to select the appropriate study for which the particular observation is being conducted (i.e., the Study ID field). Once you hit "save," the observation is now active and can be used.

Conducting an observation. At the "my observations" page you will see your newly created observation organized by description, instructor name, class name, course level, department, and date created. You will also see the following functions next to each observation: "edit info" for editing these attributes, "begin," "edit codes" for returning to your coded data for checking or alteration, and "finalize" which is clicked when you are completely done with the observation.

Start the observation by clicking on the "begin" link, which takes you to the protocol itself. You will see at the top right two sets of timers and a box indicating the current interval. The timer on the left is for the entire class session and will continue running as long as you are collecting data. The timer on the right is for the 2-minute interval only, and will refresh at the end of each interval. At that point, the interval box will also change from "1" to "2." Using these functions will help you keep track of the length of the class, your location within each interval, and the number of intervals being coded. Also note that at the end of each 2-minute interval all coded data will clear.

As you observe the class, simply click on the observed code. If you make a mistake and wish to "uncode," just click twice. To take notes, click on the box and begin typing notes regarding that interval. If taking notes, be aware of your location in the interval itself, since all data (and typing) will clear at the end of each 2-minute interval.

If you wish to cancel the observation and not store any data, hit the "cancel observation" button at the bottom of the page. The cancel function will stop the observation and erase all data. If the observation has proceeded according to plan, click on the "end observation" button when you are done. This will save all data and send you back to the "my observations" page. If you wish to check your work or alter any codes prior to "finalizing" the observation, click on "edit codes" to review your work. When you are ready to completely finish the observation, click on "finalize," which will move the observation from the active to the finalized state where no further information can be edited. These finalized observations are located at the bottom of the "my observations" page.

Downloading results and computing IRR. Administrators will need to click on "Account management" and then "Group management." At the "Current group" page you will also see the following links next to each of your groups: "edit," "study results," and "delete." The edit tab will allow you to change the name of the group, add new users, or remove existing users. The study results tab will take you to a page where all studies linked to that group are listed, and for each study you can either "download observations" or "compute statistics." If multiple studies are linked to that particular group, you will see links for each of these studies here.

The download observations link will export all of your TDOP data into a .csv file, and the compute statistics link will send you to a page to calculate IRR between pairs of users in your study. On this page, select two users (and salient observations) to compare and then click on the link for "calculate." The new page will appear to be the same as the last, but scroll to the bottom and you will see the results of your IRR test for two types of statistics: Cohen's kappa and Jaccard similarity. For each of the 6 categories in the TDOP, you can download results into .csv files to examine which specific codes are the most problematic for each pair of users.

Removing observations, studies, or groups. You may wish to delete observations, studies, or groups at some stage of the process. This is especially true in cases where you create a set of observations for IRR testing and wish to delete them prior to actual data collection. To delete individual observations, navigate to the "Studies" page under account management and click on the study being considered. Then click on the "study site" where you want to remove the observation and click on "permanently remove observation."

To remove a study, at the "Studies" page you will see a "view" function at the bottom where you can filter the studies for "all" or "posted and active." You can "delete" a study and have it removed from the active group, but these deleted studies will not be permanently removed. To permanently remove a study from the database click on the link for removing that study. We built in this extra layer in case an administrator realizes later that they did not actually want to delete a particular study. To remove a group, navigate to the "Current groups" page and click on the delete link next to the group name.

Data storage and confidentiality. All data collected via the TDOP website will be stored on secure servers at WCER. All data will be held on these servers for 1 year, so please make sure to download your data as soon as possible. At no time can anyone see the data collected for your study – only the project administrator can view data for any given study.

How to Manage, Analyze, and Interpret TDOP Data

Data collected with the TDOP can be analyzed and interpreted in a variety of ways, and in this section we describe a few of the techniques we have used in our research program to manage, analyze, and interpret the data.

Managing data: Developing a code by respondent data matrix

The first step is to convert the observation data into a data matrix that can be used for different analytic purposes. The raw data for this analysis are in the form of a two-mode (or "affiliation") matrix that consists of instructors' five-minute intervals as rows (mode 1) and TDOP codes as columns (mode 2).[5] As the table below illustrates, a '1' denotes that the particular TDOP code was present in the interval, and a '0' denotes that the code was not present in that interval.

Table 2. *Example of initial two mode matrix showing instructor-intervals (mode 1) by instructional codes (mode 2)*

Instructor	2-min. Interval	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6	Code 7
1	1	1	0	0	1	0	0	1
1	2	0	1	0	1	0	0	1
1	3	1	0	1	0	1	0	0
1	4	0	0	0	0	1	1	1
1	5	1	0	1	1	0	0	0

Using this matrix, it is then possible to analyze the data to identify simple proportions of code incidence or for more complex analyses using social network analysis software.

Analyzing data

Thus far we have analyzed TDOP data to identify code proportions and prepare affiliation graphs using social network analysis.

Code proportions. Using the raw (two-mode) dataset, it is possible to identify the prevalence of particular codes by calculating the simple proportion of 2-minute intervals in which the code was observed. This can be done in a spreadsheet by adding the number of times a particular code was observed for an individual or group (e.g., lecturing = 12 times) and dividing that number by the total number of intervals observed for the individual or group (e.g., 30). For this example, lecturing was observed in 40% of the observed 2-minute intervals. This is an easy and intuitive way to analyze the data, though the interactions between and among codes is not captured.

Table 3. *Proportion of 2-minute intervals each teaching practice was observed by discipline.*

	Entire Sample (n=58)	Math (n=14)	Biology (n=14)	Geology (n=12)	Physics (n=9)	Chemistry (n=9)
Teaching Methods						
Lecture	.21	.11	.25	.28	.20	.20
Lecture: Pre-made visuals (e.g., PowerPoint)	.48	.18	.69	.76	.35	.36
Lecture: Hand Made visuals (e.g., chalkboard)	.44	.78	.17	.18	.55	.59

5 This means that, at least initially, each instructor has multiple rows of data, one for each two-minute interval that was observed.

Lecture: Demonstration	.03	.00	.01	.01	.13	.04
Lecture: Interactive (i.e., 2+ questions posed)	.02	.02	.06	.01	.01	.01
Small group work	.09	.06	.16	.03	.13	.08
Deskwork	.06	.12	.06	.01	.07	.04
Pedagogical Strategies						
Movement	.07	.09	.07	.04	.11	.05
Humor	.07	.08	.12	.06	.03	.07
Illustration	.10	.03	.12	.15	.17	.06
Organization	.08	.07	.07	.09	.07	.11
Emphasis	.05	.04	.06	.06	.03	.08
Assessments	.05	.02	.06	.04	.08	.06

Another way to analyze the data to capture interactions between and among codes is to identify "practice triads" or other combinations of codes. A "practice triad" represents the affiliation of codes from three of the dimensions of observed practice. For example, the practice triad of "lecture-recall/retain-PowerPoint slides" can be observed in 50.7 percent of the intervals. This means that in half of the observed intervals, the teaching technique of 'lecturing' was co-coded with the cognitive engagement of 'recall/retain information' and the technology 'PowerPoint slides' in the same 2-minute interval. This technique can be done for any set of codes (e.g., pairs, triads, tetrads) and follows the same technique of identifying code co-occurrences as outlined above.

The easiest way to calculate dyads is to construct a code-by-code affiliation matrix (next section for details), and use the raw cell counts for the numerator. For more complex combinations (e.g., triads, tetrads, etc.) you must use an if/then procedure (in Excel, SPSS, or other software package) to count the number of intervals in which your condition is satisfied. The number of intervals in which your condition is satisfied becomes the numerator and the denominator remains the total number of intervals. For example, if your if/then procedure reveals that the 'lecture-recall/retain-laptop/slides' triad was satisfied in 20 of the 50 observed intervals, then this triad was observed in (20/50 = 0.40) 40% of all intervals.

Social network analysis. In some of our analyses (e.g., Hora & Ferrare, 2012) we use techniques from social network analysis to analyze the observation data, as it is well suited to capturing configurations within and between the dimensions of practice – which is at the heart of systems-of-practice theory. Using UCINET, the two-mode data matrix was transformed into a one-mode (code-by-code) affiliation matrix. This transformation results in a valued affiliation matrix in which each cell corresponds to the number of intervals code *i* is affiliated with code *j*. For example, if the intersection of 'small group work' and 'problem solving' has a value of 37, then these two dimensions of instruction are affiliated in 37 two-minute intervals across all instructors.

Next, the program Netdraw (which is included as a module in UCINET) was used to draw the valued affiliation network. The lines connecting the codes denote the strength of their affiliation. Thus, thicker lines correspond to stronger affiliations, which can be interpreted as more frequently occurring in the same two-minute interval (than those whose lines are thinner). The affiliation graphs can be used on their own as a formative evaluation instrument and will form a core element of the findings presented to local program staff. These graphs can be produced for individuals, departments, and entire institutions. In addition, the graphs can also be used to produce quantitative measures of various aspects of the graphs themselves, including the density of the entire graph, code centrality, and the proportion of times two codes co-occur. Examples of affiliation graphs are included below.

Figure 1. Affiliation graph for math instructors (n=14).

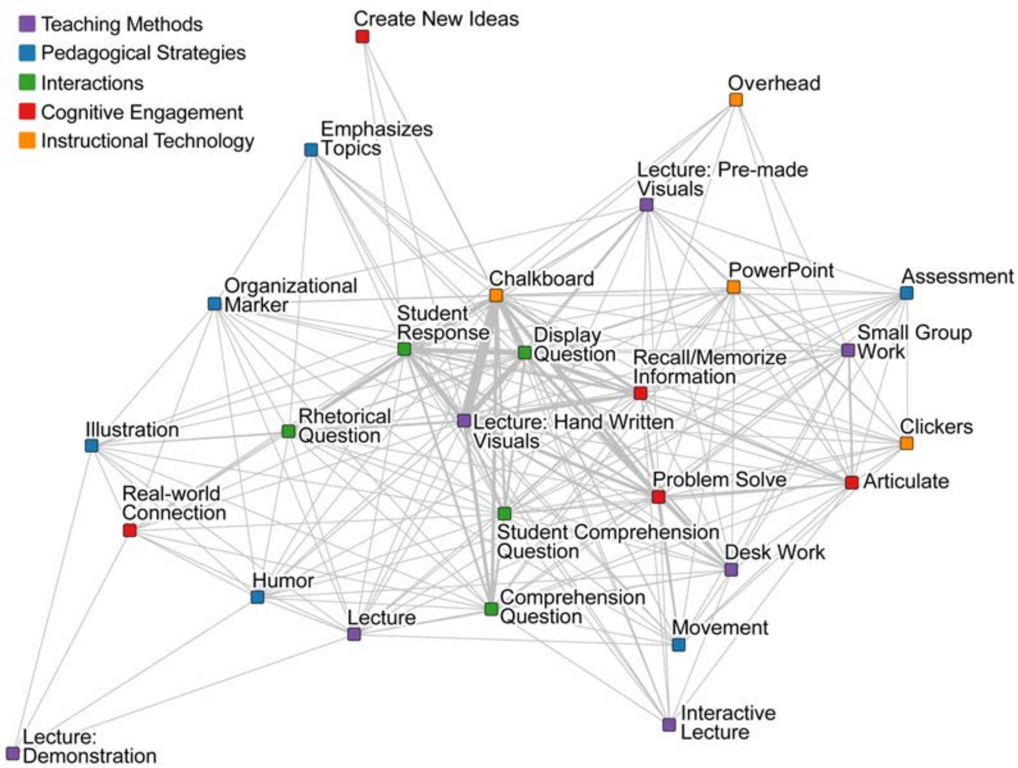
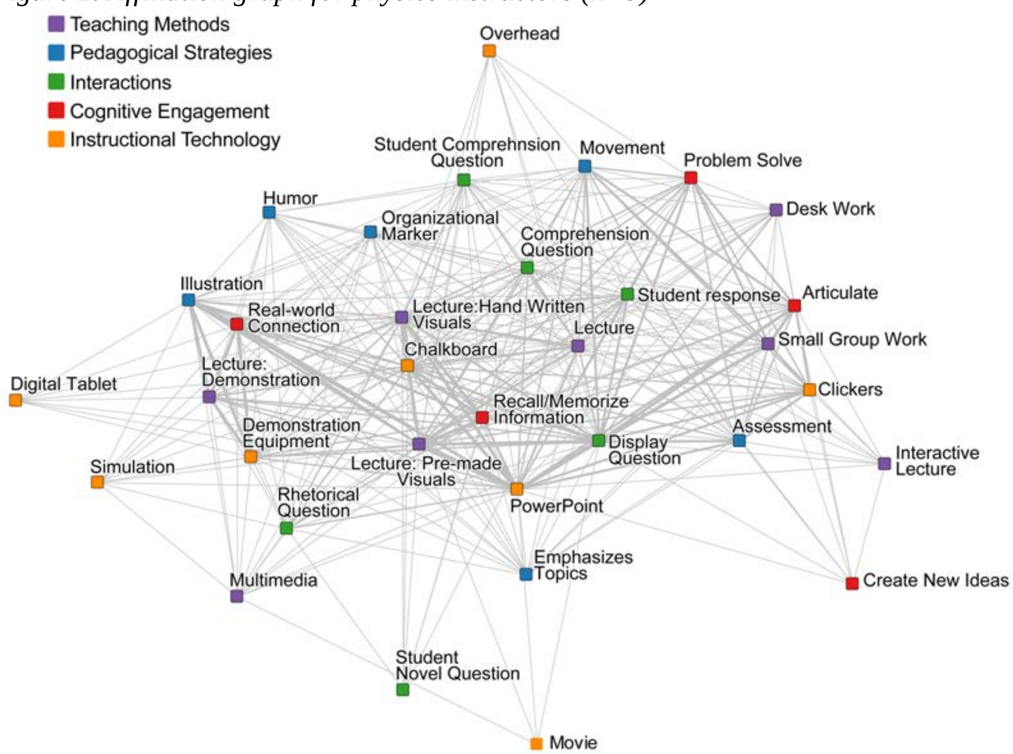


Figure 2. Affiliation graph for physics instructors (n=9)



Interpreting TDOP Data

First and foremost, we cannot emphasize enough that at the present time, TDOP data should be interpreted only as *describing* teaching practice, and not as measuring the quality of teaching and/or learning. At some point in the future it may be possible to link certain combinations of TDOP codes and/or affiliation graphs to more or less effective instruction, but that type of analysis has not yet been conducted.

In terms of describing teaching, we often interpret the data in terms of how different codes interact with one another. A multidimensional measure of teaching practice necessarily goes beyond merely capturing the different dimensions of classroom practice independently. In addition, such a measure must examine the extent to which these dimensions are interrelated. For instance, it is not only important to know how often lecturing with pre-made visuals was observed among physicists, but also how often this method was observed with the other dimensions of teaching. Figure 1 illustrates the interrelationships within and between each dimension of teaching among the mathematics instructors in our sample. The lines connecting the different TDOP nodes vary in thickness on a scale of 1 – 8 depending on the frequency with which each pair of codes was co-observed in the same 2-minute interval. Thus, thicker lines represent more frequently co-observed teaching practices. However, the positions of the nodes are not completely arbitrary. Using an iterative "spring embedding" procedure, the nodes are moved to locations in the graph that minimize the variation in line length.

From the graph in Figure 1, it can be seen that the mathematics instructors—as an aggregated group—had a core repertoire of practice that frequently made use of lecturing with handwritten visuals at the chalkboard, while posing display questions prompting student responses. The primary forms of cognitive engagement during these activities were problem solving and recalling/retaining definitions and equations. Secondarily, these instructors supplemented their core repertoire with deskwork, comprehension and rhetorical questions, and responding to student-posed conceptual questions. Less frequently, these instructors utilized pedagogical moves such as organization, emphasis, illustrations, and humor. Occasionally instructors engaged students in articulating ideas, making real-world connections, and creating, as well as lecturing with slides using PowerPoint or the overhead projector.

Figure 2 illustrates the graph for physicists, which is considerably more "dense" than the mathematics instructors' graph in Figure 1. Rather than a core/periphery structure, the physics instructors' graph reveals a wider variety of relationships within and between the different dimensions of instruction. There are three relatively distinct repertoires at work. First, on the left side of the graph, it can be seen that these instructors frequently lectured with demonstrations that involved illustrations and engaging students in real-world connections. Second, the instructors used a combination of other lecturing styles (handwritten visuals, pre-made visuals, and "straight" lecture), while engaging students in recalling/retaining information and posing display and comprehension questions that prompted student responses. Third, these physics instructors frequently used clickers to engage students in articulating ideas and problem solving in small group situations. In each case, the core practices were supplemented with additional pedagogical moves, interactions, and technologies.

To be clear, there are no rigid boundaries separating the three repertoires in Figure 2. Nevertheless, the positions of the nodes suggest a degree of distinctness to these different sets of relationships. More importantly, the graphs help to visualize disciplinary differences and the relational specificity to these differences. The graphs can also be used to examine practices among individual instructors. Indeed, such graphs are significantly easier to interpret due to the reduction of lines and nodes.

Ultimately, the data obtained through the use of the TDOP provides a detailed account of faculty teaching that opens up the "black box" of what kind of instruction happens in postsecondary settings. Since prior studies of faculty teaching have relied primarily on self-reported survey and/or interview data, while also focusing exclusively on a single dimension of teaching (i.e., teaching methods), very little detailed information is known about the types of instruction that students actually experience in the classroom. With the TDOP instrument, a more nuanced portrayal of faculty teaching is possible. For example, faculty teaching is often critiqued for relying too heavily on the "lecture," which is a common descriptor for the entirety of an individual's classroom behaviors. Yet the data reported here clearly demonstrate that teaching is actually comprised of multiple dimensions of practice and cannot be adequately described through a sole focus on teaching methods alone.

The heterogeneity of lecturing styles speaks to the more general relational framework in which these data should be interpreted.[6] That is, faculty's instructional actions in the classroom are not only comprised of teaching methods, but also strategic pedagogical moves, interactions with students, cognitive engagements, and the use of instructional technology. The regular combinations of actions within and between dimensions constitute "pedagogical repertoires" that can be examined individually or through distinct groupings (e.g., institutions, disciplines, tenure rank, etc.). Thus, a math instructor may have a repertoire that consists of the regular use of lecturing through handwritten visuals on a chalkboard. These teaching techniques and use of instructional technology may be interspersed with occasional humor and emphasis, as well as posing questions that engage students in problem solving. Another math instructor may also regularly lecture with handwritten visuals on the chalkboard, but may never pose questions or place emphasis on certain aspects of the curriculum.

Another way in which pedagogical repertoires can (and should) be examined is at the disciplinary level. Instructional behaviors such as the use of deskwork, question-posing, problem solving cognitive engagements, and chalkboards appear to be more widely used by math faculty than by others. Similarly, lecturing with PowerPoint, using illustrations, and asking students to make connections to the real world was observed more often among biology, geology, and physics faculty than other groups. These results support prior research suggesting that different disciplines have unique approaches to teaching.[7]

---

[6] Note that here we are only speaking of describing teaching-related actions in the classroom, and not the relations that influence how instructors teach.

[7] See Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories.* Open University Press; Saroyan, A. & Snell, L.S. (1995). Variations in lecturing styles. *Higher Education*, 33, 85-104; Marincovich, M. (1995). Concluding remarks: On the meaning of disciplinary differences. *New Directions for Teaching and Learning*, (64), 113-118.

Limitations to the TDOP

It is important to acknowledge limitations to the TDOP instrument. For example, since the instrument focuses on 2-minute intervals there are instances in which some codes appear to be used in combination with others when it may not necessarily be true. This is especially problematic when examining combinations of codes or comparing the proportion of instructional time across codes. Another limitation is the extensive effort required to establish inter-rater reliability. While many codes require very little training to ensure reliability, some of the pedagogical moves (e.g., emphasis and organization) and cognitive engagements proved to be quite challenging in this regard. Finally, the TDOP does not capture the "quality" of instruction. There are many instances in which different instructors use the same techniques, but with dramatically different results. In part, this is because the TDOP is focused on the instructor's observable behaviors. Yet this limitation merely suggests that the TDOP should be used in combination with other data collection tools, such as student interviews, surveys, and focus groups. More specific curricular features may also be integrated into the TDOP, but this would require a high level of content expertise. For the time being, the TDOP emphasizes those aspects of pedagogical action that can be directly observed regardless of the observer's level of familiarity with the content.

References

Amidon, E.J., & Flanders, N.A. (1967). *The role of the teacher in the classroom: A manual for understanding and improving teachers' classroom behavior.* (Rev. ed.) Minneapolis, MN: Association for Productive Teaching.

Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories*. Open University Press.

Blank, R. K., Porter, A., & Smithson, J. (2001). *New tools for analyzing teaching, curriculum and standards in mathematics and science. Results from Survey of Enacted Curriculum Project*. Final Report to the National Science Foundation on Contract No. REC98-03080. Washington, DC: Council of Chief State School Officers.

Blumenfeld, P.C., Kempler, T.M., & Krajcik, J.S., (2006). Motivation and cognitive engagement in learning environments. In R. K. Sawyer, (Ed.), *The Cambridge handbook of the Learning Sciences* (pp. 475-488). New York, NY: Cambridge University Press.

Cohen, D. K., & Ball, D. L. (1999). *Instruction, Capacity, and Improvement*. Consortium for Policy Research in Education Rep. No. RR-43. Philadelphia: University of Pennsylvania, Graduate School of Education.

Daly, A.J. (2010). *Social network theory and educational change*. Cambridge, MA: Harvard Educational Press.

Engeström, Y. (1996). Interobjectivity, ideality, and dialectics. *Mind, Culture and Activity, 3*, 259-265.

Good, T., & Brophy, J. (2000). *Looking in classrooms*. (8th Ed). New York: Longman.

Gilbert, M.B. & Haley, A. (2010). Faculty evaluations: An alternative approach based on classroom observations. *Faculty Focus*. Retrieved from http://www.facultyfocus.com/articles/faculty-evaluation/faculty-evaluations-an-alternative-approach-based-on-classroom-observations/

Gutierrez, K.D. & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher, 32* (5), 19-25.

Halverson, R. (2003). Systems of practice: how leaders use artifacts to create professional community in schools. *Educational Policy Analysis Archives*, *11* (37), 1-35.

Hora, M. T. & Ferrare, J. J. (2012). Instructional systems of practice: A multi-dimensional analysis of math and science undergraduate course planning and classroom teaching. *Journal of the Learning Sciences*, first published on September 24, 2012 as DOI: 10.1080/10508406.2012.729767.

MacIssac, D. & Falconer, K. (2002). Reforming physics instruction via RTOP. *The Physics Teacher, 40,* 479-485.

Marincovich, M. (1995). Concluding remarks: On the meaning of disciplinary differences. *New Directions for Teaching and Learning,* (64), 113-118.

Morell, T. (2004). Interactive lecture discourse for university EFL students. *English for specific purposes, 23*(3), 325-338.

Murray, H.G. (1999). Low-inference teaching behaviors and college teaching effectiveness: Recent developments and controversies. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research*. Volume 15. New York: Agathon.

Murray, H.G. (1983). Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. *Journal of Educational Psychology*, 75, 138-149.

Nystrand, M. & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English, 25* (3), 261-290.

Osthoff, E., Clune, W., Ferrare, J., Kretchmar, K., and White, P. (2009). *Implementing Immersion: Design, professional development, classroom enactment and learning effects of an extended science inquiry unit in an urban district.* Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Educational Research.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher 31*(7), 3-14.

Saroyan, A. & Snell, L.S. (1995). Variations in lecturing styles. *Higher Education*, *33*, 85-104.
Turpen, C. & Finkelstein, N.D. (2009). Not all interactive engagement is the same: Variations in physics professors' implementation of peer instruction. *Physical Review Special Topics – Physics Education Research, 5* (2), 1-18.
Wainwright, C.L., Flick, L.B., & Morrell, P.D. (2003). Development of instruments for assessment of instructional practices in standards-based teaching. *Journal of Mathematics and Science: Collaborative Explorations, 6* (1), 21-46.