Exploring the Use of the Teaching Dimensions Observation Protocol to Develop Fine-grained Measures of Interactive Teaching in Undergraduate Science Classrooms

WCER Working Paper No. 2013-6

October 2013

Matthew T. Hora
Assistant Scientist
Wisconsin Center for Education Research and the Center for Education and Work
hora@wisc.edu

Wisconsin Center for Education Research
School of Education • University of Wisconsin–Madison • http://www.wcer.wisc.edu/

Hora, M. T. (2013). *Exploring the Use of the Teaching Dimensions Observation Protocol to Develop Fine-grained Measures of Interactive Teaching in Undergraduate Science Classrooms* (WCER Working Paper 2013-6). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/workingPapers/papers.php

Exploring the Use of the Teaching Dimensions Observation Protocol to Develop Fine-grained Measures of Interactive Teaching in Undergraduate Science Classrooms

Matthew T. Hora

What do we know about how science is taught in today's undergraduate classrooms? Currently, data on teaching in college and university classrooms is generally limited to end-of-term student evaluations, intermittent and unstructured peer observations, and self-reported questionnaires. While some surveys report that 63% of faculty[1] report "extensive lecturing" as their primary mode of instruction (Hurtado et al, 2012), in asking respondents to self-report their teaching practices across multiple courses, the results necessarily reflect a coarse measure of what happens in a teacher's classroom on a day-to-day basis. Further limiting the field of postsecondary and science education's ability to produce rigorous and fine-grained accounts of teaching is the reliance on a binary categorization scheme to classify instruction—that of lecturing and interactive teaching (e.g., Freeman et al, 2014). Lecturing as a mode of verbal exposition can: (1) be characterized along multiple dimensions including organization and degree of student interactions (Perry & Smart, 1997), (2) vary in intensity or "dosage," and that (3) brief, well-designed periods of lecturing can be pedagogically effective (Saroyan & Snell, 1997). As a result, beyond accounts of the pervasiveness of lecturing obtained from self-reported surveys, we actually know very little about what happening in today's undergraduate science classrooms - a problem that is compounded by the lack of research instruments that can reliably capture instructional practice at a fine-grained level.

Why does this matter? First, considerable resources are being devoted to improving undergraduate science education, and these improvements largely focus on altering the pedagogical approaches faculty use in the classroom from the "sage on the stage" model to one that more directly engages students in their own learning (e.g., President's Council of Advisors on Science and Technology, 2012). I use the generic term "interactive teaching" in this paper when referring to those types of approaches. In response, researchers and policymakers alike are interested in assessing the degree to which interactive modalities are being adopted in order to evaluate the impacts of these investments and the underlying processes of educational change (Henderson, Beach & Finkelstein, 2011). Second, research on educational reform suggests that local populations are more receptive to new policies when they are closely aligned with existing cultural norms and practices (Spillane, Reiser & Reimer, 2002). Given evidence that faculty are reacting negatively to top-down reforms that ignore local knowledge (Henderson & Dancy, 2008), as well as the potential for lecturing to be an effective pedagogical device (Small, 2014), it is essential for the field to develop interventions that build upon, rather than propose to eliminate, the types of educational practices such as lecturing that are currently in use.

Increasingly, to understand the current state of teaching while also providing detailed feedback to support professional development, both K-12 and postsecondary researchers rely on classroom observations (Pianta & Hamre, 2009). The development of observation protocols is

[1] By *faculty* I mean all people who hold undergraduate teaching positions—whether full- or part-time, in a tenure track or not—in postsecondary institutions with the exception of graduate teaching assistants.

more advanced in K-12 settings due in part to policies governing teacher evaluation, with highly structured instruments such as the Framework for Teaching (Danielson, 2013) being utilized by districts across the United States. In postsecondary institutions, observation protocols have traditionally been less structured where observers take notes about their impressions of a given class, yielding rich but idiosyncratic data dependent upon the observer (Chism, 2007).

In response, my research program created the Teaching Dimensions Observation Protocol (TDOP) to be used in postsecondary settings—an adaptation of a tool developed to study middle school science teaching (Osthoff et al, 2008). The TDOP was field-tested in a three year study of teaching in the science, technology, engineering, and mathematics (STEM) disciplines (Hora & Ferrare, 2013). Using a time-sampling framework, the TDOP entails scoring the presence or absence of 47 codes across five dimensions of instruction at two-minute intervals, which results in fine-grained and standardized accounts of teaching. While users have appreciated the detailed and non-evaluative nature of the TDOP, questions remained about what precisely the results could tell the science education community about the degree to which faculty were adopting (or not) interactive teaching methods.

In this paper, I describe new techniques for combining TDOP codes to provide evidence for the presence of interactive teaching and how the temporal fluctuation of teaching can shed additional light on classroom practice. This latter focus on time is based on the widely held notion that because of students' limited attention spans (i.e., 15-20 minutes) the duration of class activities should be correspondingly limited (e.g., Middendorf & Kalish, 1996), as well as the need for the field to determine the appropriate "dosage" or time spent on lecturing and interactive modalities (Freeman et al, 2014). Using analytic and data visualization techniques such as social network analysis and time-series graphs, I illustrate the utility of the TDOP by reporting findings from a new study that involved observations of 56 instructors in biology, mechanical engineering, geology, and physics departments to address three research questions:

(1) What are the classroom practices employed by a group of science and engineering faculty?

(2) To what degree are faculty using interactive teaching practices?

(3) How, if at all, can time-sampled data be used to provide insights about teaching?

The results indicate that various forms of verbal exposition, particularly that of lecturing with pre-made visuals (e.g., PowerPoint slides), are a dominant form of instruction among the study sample (observed in 64% of all two-minute intervals), though the use of other behaviors such as different pedagogical strategies, student-teacher interactions, and instructional technologies can result in two classes with heavy PowerPoint use looking rather different in practice. Additionally, the data indicate that students are spending an appreciable time engaged in interactive teaching modalities, including answering verbal questions (19%), electronic questions via clickers (8%), engagement in desk work (16%), and participating in group-based work (11%). Analyses focused on the temporal nature of teaching yield further insights, including the fact that 61% of the faculty lectured for no longer than 20 minutes within a given class. The results indicate that "lecturing" is an imprecise term that masks a variety of instructional behaviors when its multiple forms and dosage are not taken into account.

Ultimately, I argue that fine-grained observation data obtained using the TDOP that provide robust, non-evaluative insights about teaching can be used to assess current practice, support the design of locally attuned interventions, and inform the development of research that moves beyond the lecturing versus interactive teaching dichotomy, or what Freeman et al. (2014) call "second-generation research" on undergraduate teaching.

Background

In this section I briefly review the literature on postsecondary and STEM teaching, methodological issues with the study of teaching, and background of the TDOP.

Current Knowledge on Postsecondary and STEM Teaching

In attempting to characterize different aspects of teaching, researchers have adopted a variety of approaches. One line of inquiry that set the stage for much of the current work on STEM education focused on discerning differences in the pedagogical thoughts of faculty such as beliefs about teaching and learning (Samuelowicz & Bain, 2001) and approaches to teaching (Hativa & Goodyear, 2001). This body of research concluded that the thoughts of postsecondary teachers could be characterized as having an underlying dimensionality that exists on a continuum from student-centered (where teaching is seen as the active facilitation of student learning) to teacher-centered (where teaching is seen as the delivery of content) (Kember, 1997). However, this body of research has been critiqued for its reliance on self-reported survey data as the basis for establishing links between cognition and action (Kane, Sandretto & Heath, 2002), and that "a strong opposite 'either/or' positioning of the approaches does not do justice to the nature of the phenomenon" (Postareff & Lindblom-Ylanne, 2008, p.120).

Another approach focuses on classroom practice itself, and the strategies that can be linked to student learning (see Perry & Smart, 1997 for a review). One of the key insights from this literature is that besides teaching methods such as lecturing or small group work, instructors use strategies such as humor, anecdotes, and verbally marking transitions between topics (i.e., organization) that play critical roles in the quality of instruction (Feldman, 1989; Murray, 1983). Interestingly, a body of literature also exists on ways to characterize the oft-used lecturing method with researchers emphasizing that lecturing as a form of verbal exposition varies along multiple dimensions that include organization and the degree of interactiveness with students (McKeachie, 1994), personal lecturing styles (Brown & Bakhtar, 1998), and the degree to which considerations of content, context, or pedagogy ultimately shape the lecture and its delivery (Saroyan & Snell, 1997). While each of these studies focus on some form of instructor verbal exposition, some argue that a major impediment to the field is the lack of an "operational definition of the classic lecture" (Schonwetter, 1993, p.2).

One aspect of teaching that has garnered considerable interest is the degree to which students are actively engaged in the learning process, which is largely due to the considerable evidence generated on the topic from research on learning and human development (e.g., Bransford, Brown & Cocking, 1999). Student engagement is a multi-dimensional construct that encompasses cognitive, emotional, and behavioral engagement, and in the classroom can take many forms including the posing of questions and peer-to-peer engagement (Fredricks, Blumenfeld & Paris, 2004). Many pedagogical strategies have been developed to foster increased student engagement in the classroom, such as problem-based learning (PBL) where students work collaboratively in groups on real-world problems (Hmelo-Silver, 2004).

Increasingly, the STEM disciplines have been the focus of educational reforms promoting student engagement in the classroom. One such strategy is Peer Instruction, which engages students in activities that require applying concepts to problem-solving situations and then discussing solutions with peers (Crouch & Mazur, 2001). Another widely used strategy is that of Scientific Teaching, which "involves active learning strategies to engage students in the process of science" (Handelsman et al., 2004, p. 521), while also advocating for the use of teaching methods (e.g., problem-based learning) whose effectiveness has been determined with the same rigor that scientists use in their own disciplinary research. As previously noted, the widespread dissemination of classroom practices such as Peer Instruction and Scientific Teaching are a priority of policymakers in the U.S. and abroad (e.g., PCAST, 2012).

But two important qualifications exist to this focus on reforming how faculty teach their classes. First, researchers are also exploring how additional aspects of instructional design can facilitate student learning, such as requiring students to complete pre-class reading quizzes (Crouch & Mazur, 2001; Freeman, Haak & Wenderoth, 2011). This focus on out-of-class student work implicitly acknowledges that while teachers play a critical role in designing effective learning environments, they are not the only actor responsible for the learning process. Second, the utilization of interactive teaching strategies alone does not guarantee improved learning experiences for students. For example, Turpen and Finkelstein (2009) found that some physics instructors failed to use Peer Instruction techniques the way they were intended, with the results being pedagogically ineffective learning experiences for students.

Research on the Prevalence of Classroom Practices

In contrast to research that examines the efficacy of specific pedagogical techniques or underlying principles of teaching, less research exists that surveys the types of teaching practices being used by faculty in the field. In perhaps the most comprehensive analysis of faculty teaching in the U.S., the Higher Education Research Institute (HERI) conducts a survey of faculty work that includes items on teaching, advising, and job satisfaction. In the 2010-2011 survey, which included 23,824 full time faculty at 417 postsecondary institutions across the U.S., researchers found that STEM faculty (n=6,768) reporting using "extensive lecturing" in 63% of the courses they teach, "class discussion" in 61.5%, "cooperative learning (small groups) in 47%, and "using student inquiry to drive learning" in 36.5% of their courses (Hurtado et al, 2012).

Given the substantial investments in encouraging STEM faculty to adopt interactive teaching methods, researchers are also conducting survey-based studies to assess the degree to which these innovations are being used. For example, a survey of 722 physics instructors found that 29% used Peer Instruction, 13.9% used interactive lectures, and 13.7% used cooperative group problem-solving (Henderson & Dancy, 2009). Another survey of 99 electrical engineering faculty found that 60% used active learning methods, 35% problem-based learning, and 15% Peer Instruction (Prince et al, 2013). Given the logistical issues with conducting classroom observations, it is unsurprising that the literature contains few instances in which observations have been used to assess the nature of teaching in undergraduate STEM courses on a broad scale.

Methodological issues with the study of postsecondary teaching. In addition to the relative paucity of evidence on instructional practice, methodological issues related to research instruments and design further inhibit the field.

Challenges with self-report surveys. In a critical review of validity problems associated with surveys of college students, Porter (2011) highlighted issues with the ability of students to accurately and reliably recall their experiences over imprecise periods of time (i.e., response process). For surveys such as the HERI Faculty Survey (Hurtado et al, 2012) that asks faculty to report their teaching across “most classes,” a similar concern about the process of respondent recall and reporting exists. Additionally, little research exists on the relationship between self-reported teaching behaviors and actual classroom practices, such that assuming that reported practice reflects actual practice is problematic (Mayer, 1999; Kane et al, 2002). Other limitations with some survey studies include low response rates (e.g., 13.2%) and that the effectiveness with which a reported practice is implemented in the classroom is unknown (Froyd, et al, 2013). Finally, survey instruments tend to focus on a single dimension of practice (i.e., teaching methods), instead of the variety of dimensions known to constitute effective teaching, such as instructor clarity (Feldman, 1989) while also failing to account for the intensity of particular methods used within a class period (e.g., whether lecturing is used in 10% or 100% of a class).

Pervasive use of the construct of “lecturing.” Further complicating the scientific measurement of teaching is the tendency to reduce the complex phenomenon that is teaching into single descriptors such as “lecturing,” One reason this is problematic is the lack of operational definitions that describes precisely what types of instructional practices are encompassed within a lecture (Schonwetter, 1993; Saroyan & Snell, 1997). For example, in a widely-cited educational experiment comparing lecturing to interactive teaching, the interactive condition is described as follows: “There was no formal lecturing; however, guidance and explanations were provided by the instructor throughout the class” (Deslauriers, Schelew & Wieman, 2011, p.863). The precise differences between “guidance and explanations” and “formal lecturing,” as well as the percentage of class time was devoted to any particular approach, is not provided in the paper. Besides failing to articulate the precise nature of the treatment and control conditions in experimental studies, which raises questions about the validity of the results (Derting et al, 2011), using blunt descriptors such as lecturing also mask features of instruction that may be subsumed within a period of verbal exposition. That is, within a “lecture” an instructor may utilize principles of effective instruction such as organization (Perry & Smart, 1997), intersperse the discussion with questions, or use the lecture to set up interactive modalities.

Unfortunately, the reliance on descriptors such as lecturing is reinforced by their inclusion in instruments such as the HERI Faculty Survey (Hurtado et al, 2012), which codify such reductionist conceptions of teaching. In addition, except for observation protocols that utilize time-sampling approaches, the intensity of teaching practices is left unmeasured. Thus, it is clear that if the field is to begin to measure and compare instruction at a more fine-grained level than what is captured by descriptors such as lecturing, new instruments will be required.

A New Approach: Using Classroom Observations to Study Postsecondary Teaching

In response to questions about the ability of survey data to adequately describe classroom activities, researchers are increasingly turning to classroom observations to obtain firsthand data on teaching practices (Pianta & Hamre, 2009). Examples of widely used instruments in postsecondary settings include the Teaching Behaviors Inventory (Murray, 1983) and the Reformed Teaching Observation Protocol (RTOP) (MacIsaac & Falconer, 2002). The RTOP in particular has been widely used to evaluate the degree to which classroom instruction is aligned with national standards in reform-based teaching (e.g., Ebert-May et al., 2011). In practice, analysts observe a class while taking un-structured notes, and at the conclusion of the lesson answer 25 Likert-style questions along dimensions such as lesson design and implementation and classroom culture. However, considerable questions remain about the reliability of instruments that determine instructional quality a priori. A recent review of the reliability of evaluative protocols used in the Measuring Effective Teaching Project found that ratings varied considerably across observers regardless of training (Guarino & Tracy, 2012), and that rater bias (i.e., pre-existing beliefs about what constitutes high-quality teaching) is a major reason for the high degree of variability observed in the use of these protocols (Cash et al, 2012).

The TDOP. Partly in response to these issues, the TDOP was developed as a descriptive rather than an evaluative instrument. Instead, the aim was to produce rich, fine-grained accounts of teaching that allowed for the measurement of multiple dimensions of instructional practice over time. As such, the appropriate unit of analysis for behavior is not the individual but larger activity systems that encompasses individuals and their situations in an integrated whole (Cole, 1996). 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" in educational settings. With this theoretical stance in mind, my research group adapted an observation protocol designed for studying interactive instruction in middle school classrooms (Osthoff et al., 2008) for use in postsecondary settings. The TDOP is comprised of five categories posited to represent critical dimensions of instruction that not only includes the use of specific *teaching methods* (e.g., small-group discussion), but also the types of *pedagogical strategies* they use in the classroom (e.g., organization), the types of *student-teacher interactions* in the classroom (e.g., types of questions posed), the *potential cognitive engagement* that teachers demand of 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). Each category contains several codes—though it is important to note that the categories do not represent latent constructs but instead are simply groups of codes that capture distinct aspects of teaching. For a description of the codes that comprise each category see Table 1[2].

[2] This list of TDOP codes reflects those codes included in the analysis reported in this paper. A total of 47 codes were included in the version of the protocol used to collect the data, but in the interest of reporting an accessible dataset I eliminated 14 codes from the dataset for use in this paper. The complete protocol is available at http://tdop.wceruw.org/.

***Table 1*. Description of TDOP categories and codes**

TDOP Categories	Codes	Description of Codes
Teaching Methods		
Lecturing	L	Instructor speaks to students w/no media
Lecturing w/pre-made visuals	LPV	Instructor speaks w/pre-made visual media (e.g., PowerPoint slides)
Lecturing w/hand-made visuals	LHV	Instructor speaks to students w/hand-made visuals (e.g., writing on chalkboard)
Lecturing w/demonstration	LDEM	Instructor speaks while using demonstrations
Socratic lecture	SOC-L	Instructor speaks while asking questions (2 or more), the answers to which guide the discussion.
Working through problems	WP	Instructor works out computations or problems.
Small group work	SGW	Students form into groups of 2+
Desk work	DW	Students complete work alone at desk
Multi-media	MM	Instructor plays a video/movie without speaking
Assessment	A	Instructor gathers student learning data
Pedagogical Moves		
Humor	HUM	Instructor tells jokes (2+ students must laugh)
Anecdote/example	ANEX	Examples that link material to student experiences
Graphic	GR	Instructor uses graphic image to illustrate material
Organization	ORG	Instructor clearly indicates transition between topics
Emphasis	EMP	Instructor clearly states a topic is important
Teacher-Student Interactions		
Rhetorical questions	IRQ	Instructor poses questions w/o waiting for answer
Display questions	IDQ	Instructor poses questions seeking information
Comprehension questions	ICQ	Instructor poses question about student understanding
Student novel question	SNQ	Student asks original question
Student comprehension question	SCQ	Student asks for clarification about previous topic
Student response	SR	Student responds to instructor question
Student peer interactions	PI	Students interact with one another
Cognitive Engagement		
Problem-solving	PS	Students are asked to actively solve a closed-ended problem with a known solution
Creating	CR	Students are asked to actively solve an open-ended problem w/o a known solution
Connecting to real-world	CN	Students are given examples linking material to common experiences
Instructional Technology		
Chalkboard	CB	Chalkboard or whiteboard used for writing
Overhead projector	OP	Machine used to project images on screen
PowerPoint	PP	Microsoft PowerPoint slides
Clickers	CL	Clicker response systems
Demonstrations	D	Laboratory demonstration equipment
Digital tablet	DT	Machine used to project images and writing on screen
Movies	M	Movies (e.g., YouTube movies)
Simulations	SI	Graphic simulations and animations

Each category (and code) is meant to be viewed in interaction with the others as an integrated system of practice. The configurations that form through practice can be empirically studied through techniques such as social network analysis, which is increasingly being used to study complex dynamics within educational settings (e.g., Grunspan, Wiggins & Goodreau, 2014).

What is the validity evidence for the TDOP? Traditionally, validity in higher education has focused on establishing criterion-validity (how well a score predicts or estimates a measure that is external to the test), and construct-validity (how well a measure adequately captures the domain of interest) for surveys. Increasingly, scholars have adopted argument-based approaches to validity, which entails collecting varied sources of evidence and theory to support the interpretation of particular measures in light of their intended uses (Porter, 2011; Kane, 2001). Given the intended use of the TDOP to provide descriptive accounts of teaching and not to ascertain the presence of an external criterion, testing for criterion-validity was not appropriate. Instead, face and construct validity for each of the codes and categories was tested through preliminary fieldwork and feedback from disciplinary and education experts. These groups of faculty confirmed that the codes included in the instrument were consistent with their own understanding of teaching. Additionally, since groups of codes are not intended to measure latent constructs, construct validity tests on this point were not applicable.

Perhaps most importantly for observation instruments being used by multiple raters is inter-rater reliability (IRR), which ensures that different analysts will use an instrument in a similar manner across cases. The training procedure for the TDOP, which is described in the methods section, is rather extensive and places considerable focus on IRR. As further development with the TDOP continues, additional validity and reliability evidence will be gathered (e.g., test-retest reliability), though for interpretations that do not extend far beyond the observations upon which they are based (as is the case with the TDOP), extensive validity evidence is not required (Kane, 2001).

The TDOP is available in hardcopy and web-based formats, and at the time of writing a total of 233 administrators (i.e., managers of user groups) were registered on the TDOP website. The instrument has been used in a variety of research projects and has also been adapted for local uses, often with the overall structure and coding procedures intact but with codes removed and/or re-named (see Smith et al, 2013 for a minor adaptation).

Analyzing TDOP data to study interactive teaching. As previously noted, while users and the research community have generally appreciated the non-judgmental and fine-grained data produced by the TDOP, a question has been "so what?" particularly in regard to whether the data can shed light on the prevalence of interactive teaching. Towards that end, in this paper I describe six new code combinations that capture behaviors that are closely aligned with principles of interactive teaching. To operationalize interactive teaching practices in terms of TDOP codes, I referred to distinct pedagogical approaches such as Scientific Teaching (Handelsman et al., 2004) and Peer Instruction (Mazur, 1997) in order to identify observable features of instruction that were central to these approaches. From this review I distilled a core set of pedagogical principles that center on instructors actively facilitating students' engagement with the people around them, with the instructor him or herself, and with the course material. I recognize that these principles do not reflect the entirety of the theories underlying Scientific

Teaching or Peer Instruction as well as their practical application in the classroom, but they do reflect a core underlying idea of active student engagement that is central to these teaching strategies. It is also important to note that these code combinations are not intended to be latent constructs that indicate the presence (or absence) of a particular aspect of interactive teaching, but instead are best viewed as one of several potential indicators for a particular type of teaching. Using observations to capture aspects of active student engagement has a long history, including time-sampling techniques where observers record when a type of behavior is present or absent (Lee & Brophy, 1996). In this paper I build on these approaches by defining student engagement in terms of TDOP codes (see Table 2).

***Table 2*. Description of TDOP code combinations used to capture elements of interactive teaching**

	TDOP Codes	Description
Type of Instruction Based on Primary Actor in Classroom		
Teacher-centered	L or LPV or LHV or LDEM or SOC-L or A	Codes where teacher is the speaker and primary locus of activity
Student-centered	DW or SGW or SQ or SR or PI	Codes where students are speaking and/or the primary locus of activity
Other Indicators		
Q&A sequence (verbal)	DQ and SR	Instructor verbally poses a question, followed by student verbal response
Q&A sequence (electronic)	DQ and CL and SR	Instructor poses a question that involves student responses using clickers
Student activity (solo)	DW or A	Students visibly engaged in completing a task on their own
Student activity (group)	SGW	Students visibly engaged in completing a task with at least one other peer

Teacher- or student-centered instruction. First, engagement can be coarsely measured by capturing whether students are visibly active in the classroom in any sort of activity (e.g., talking to one another) or not. When compared with the amount of time that students are visibly inactive and the instructor is the primary actor or interlocutor in the classroom, it becomes possible to obtain two measures of student engagement that refer to who the primary actor (the student or the teacher) is in the classroom.

Question-and-answer exchanges (verbal). Second, cases where a question is verbally posed by an instructor to his or her students, with the clear expectation that an answer is forthcoming (i.e., not a rhetorical question), represent a more fine-grained indicator of engagement. In these instances, students are actively involved in formulating a response and providing an answer to the instructor.

Question-and-answer exchanges (electronic). Third, one of the limitations with verbally posed questions is that it is possible that only a few highly motivated students will regularly interact with the instructor (Crouch & Mazur, 2001). As a result, electronic forms of question-and-answer exchanges, which typically involve clicker response systems, are also included as a fine-grained indicator of student engagement.

Student work (in groups). Fourth, one of the most common features of interactive teaching strategies is that of group work, where two or more students are given a task or problem to solve together. This approach is based on research indicating that students learn through the co-construction of knowledge with one another (Crouch & Mazur, 2001; Hmelo-Silver, 2004).

Student work (solo). Finally, student-led activity can also be conducted at the individual level in what is commonly known as "desk work." In these cases, students are directly and visibly engaged with the material at their own desks or tables.

It is important to note the limitations with this approach of utilizing TDOP codes to identify aspects of interactive teaching, which center on the fact that the degree to which a particular behavior is effective or not cannot captured. Thus, the analysis reported in this paper simply indicates the presence of an interactive teaching modality and not its actual efficacy in facilitating student learning. This means that certain results such as a low incidence of student-centered activity should not be automatically equated with ineffective teaching.

Methods

This study took place at three large, public research universities in the U.S. and Canada in the spring of 2013. Research universities were selected for this study in part because of the large number of undergraduates being trained in STEM disciplines at these institutions. The three study sites had similar undergraduate enrollments, external research funding, and were selected for inclusion in this study due to STEM pedagogical improvement initiatives underway at the time of data collection. Personnel active in these initiatives provided initial contacts for our team of educational researchers. Faculty were included in the study population if they were listed as course instructors in course listings for the spring semester. 165 individuals were contacted via email with a request to participate in the study, and 56 ultimately participated in the study (34% response rate). The participants represented the following disciplinary groups: biology (n=18), mechanical engineering (n=12), geology (n=15) and physics (n=11). These disciplines were selected due to the large populations of instructors across the study sites and for their leadership in STEM education initiatives[3]. Faculty self-selected into the study, and thus the results should not be generalized to the larger population of instructors at these institutions or in higher education (see Table 3).

3 Given that these disciplines reflect only a few of fields captured within the acronym of "STEM," for the remainder of the paper I refer to science and engineering disciplines.

Table 3. **Description of sample**

	Participants n	Percentage %
Total	56	100
Sex		
Female	18	32
Male	38	68
Discipline		
Biology	18	32
Mechanical Engineering	12	21
Geoscience	15	27
Physics	11	20
Level of course		
Lower division	34	61
Upper division	22	39
Size of course		
25 or less	8	14
26-100	18	32
101-199	18	32
200 or more	12	22
Position type		
Lecturer/Instructor	21	38
Assistant Professor	9	16
Associate Professor	14	25
Professor	12	21

The course component of interest in this study was the class period, colloquially known as the "lecture" period. That is, laboratory and discussion sections were not observed. Thirty-four lower-division and 22 upper-division courses were included in the study, the designation of which was determined by consulting each institution's course numbering system (e.g., lower-division courses at one institution were numbered 1000-2000, and upper-division courses 3000-4000). The courses also varied by enrollment numbers, which were obtained from the instructor.

Data Collection

A team of four researchers collected data at the study sites during weeklong field visits in the spring of 2013. Prior to gathering data the four researchers participated in a rigorous training program that included in-depth discussions about the meaning of each code category and individual codes, practice coding of videotaped class segments, and finally, the coding of entire videotaped lectures. This training process took two weeks, and at the conclusion of the training the team coded two full classes (i.e., one each in physics and biology) as part of IRR testing. IRR was calculated using Cohen's kappa scores averaged across two classes. [4] Cohen's kappa is an

4 It should be noted that the claim made in the Smith et al. (2013) paper that adequate training to establish IRR can be conducted in 1.5 hours runs counter to the experiences of the TDOP research group.

index that measures the level of agreement between two sets of dichotomous ratings, while taking into account the possibility that agreement can take place by chance (see Table 4).

***Table 4*. Description of TDOP inter-rater reliability scores for analysts**

	Teaching Methods	Pedagogical Moves	Interactions	Cognitive Engagement	Instruct. Technol.
Analyst 1/Analyst 2	.90	.85	.83	.74	.94
Analyst 1/Analyst 3	.82	.81	.73	.78	.90
Analyst 1/Analyst 4	.89	.74	.79	.71	.90
Analyst 2/Analyst 3	.83	.80	.81	.75	.89
Analyst 2/Analyst 4	.84	.75	.79	.77	.89
Analyst 3/Analyst 4	.80	.73	.72	.74	.91

After obtaining permission from the instructor, analysts sat near the back of the classroom in order to obtain a clear view of the entire room, and then proceeded to observe the entire class period. A total of 95 class periods were observed, with 39 faculty observed twice and 17 faculty observed once. This discrepancy was due to scheduling issues such as exam dates and courses that met only one time a week.

The study team used the online form of the TDOP that entails clicking on a code when it is observed during a given two-minute interval. It is important to note that because a variety of practices may occur within a single interval, more than one code for a given dimension (e.g., teaching methods) may be coded within the same interval. Further, in instances where a behavior started in one interval (e.g., 2:00-3:59) and ended in another (e.g., 4:00-5:59), it was coded in both intervals. While this coding procedure may seem overwhelming, with adequate training the coding scheme and corresponding cognitive load on the observer is not an issue. That being said, the challenges inherent in using a time-sampling protocol underscore the importance of rigorous training. Finally, for the IRR training and the fieldwork the team used a form of the TDOP that included 47 codes, but in this paper I report data using 33 codes in order to report a manageable set of results.

Data Analysis

Data from the TDOP instrument were exported from the online server into spreadsheets where individual two-minute intervals are rows and codes are columns. A code observed by the analyst is indicated by a "1" and not observed is represented by a "0."

Identifying frequencies for individual codes. For the set of results the proportion of times that a particular code (e.g., small group work) was observed across all two-minute intervals is reported. These data are reported for the entire sample as well as by groups that may be of interest to readers including groups by discipline, course level, and class size. It is important to note that a code is scored as present if the corresponding practice is observed for any portion of a given two minute interval. Thus, the frequencies reported reflect the portion of intervals in which the code was observed, but only roughly approximate the amount of actual class time in which the code occurred. Furthermore, since multiple codes can occur simultaneously, the sum of the various interval codes typically exceeds the total amount of class time.

Social network analysis. In addition to the frequency of individual codes and code groups, the nature of the interactions among individual codes can reveal important nuances of classroom practice. To do this I used techniques from social network analysis to delineate configurations within and between the dimensions of practice. The raw data for these analyses are in the form of two-mode (or "affiliation") matrices consisting of instructors' two-minute intervals as rows (mode 1) and TDOP codes as columns (mode 2). Using UCINET (Borgatti, Everett & Freeman, 2002) the two-mode data matrix was transformed into a one-mode (code-by-code) matrix through matrix multiplication. This transformation results in a valued co-occurrence matrix in which each cell corresponds to the number of intervals in which two given TDOP codes are affiliated. For example, the intersection of the codes for small group work and problem-solving could have a value of three, which means these two dimensions of instruction were co-coded in three intervals across all instructors in the matrix. Using these data, the program Netdraw was used to graph the co-occurrences between each pair of codes across all instructor-intervals.

Identifying frequencies for interactive teaching. Next, to identify the frequency with which interactive teaching was observed, I created new variables using SPSS statistical analysis software based on the code combinations described in Table 2. It is important to note that some new variables such as teacher-centered instruction utilized the "or" operator, such that if any of the six codes included in this category was observed, then the cell for that interval would be coded as "1." In contrast, other categories such as verbal questioning were calculated using the "and" operator, which required two or more codes to be observed within the same interval.

Depicting temporal data and calculating dosage. Finally, data from a single observation was used to create a graph that showed the presence or absence of certain codes for each two-minute interval. An observation that included a combination of both teacher-centered and student-centered modalities was selected for illustrative purposes. The class featured in this graphic is also featured in Figure 1 as "Biology Instructor #2." In addition, certain time-sensitive metrics that could be used to characterize the teaching of this instructor were calculated including the percentage of time spent lecturing, the longest period spent lecturing, and the total number of distinct learning activities. Finally, given the interest in the extent of lecturing within classrooms, I then calculated for the entire sample the longest periods each instructor spent in a lecturing modality (e.g., L, LHV, LPV and LDEM) with no observable engagement with students as measured by the following codes: LINT, CD, SGW, DW, DQ, A, SR, and PI.

Results

1. What are the classroom practices employed by a group of science and engineering faculty?

To answer the first research question pertaining to the types of teaching practices observed in the field, I first report results for each individual TDOP code for the entire sample and then by discipline, course level, and class size. Then to illustrate the inter-connected nature of distinct teaching dimensions as suggested by systems-of-practice theory, a pair of graphs created using social network analysis techniques are presented.

Table 5. **Classroom observation data using the TDOP by discipline**

Discipline	Code	All	Biology	Mech. Eng.	Geo-science	Physics
Instructors		56	18	12	15	11
Total 2-minute Intervals		2,520	751	527	767	427
Teaching Methods						
Lecturing	L	.06	.02	.06	.10	.08
Lecturing w/premade visual	LPV	.64	.86	.56	.63	.43
Lecturing w/handmade visual	LHV	.27	.11	.52	.14	.49
Lecturing w/demonstration	LDEM	.03	.02	.07	.01	.05
Socratic lecture[a]	SOC-L	.03	.06	.04	.02	.01
Working through problems	WP	.08	.02	.26	.00	.19
Small group work	SGW	.11	.11	.10	.13	.11
Desk work	DW	.07	.03	.12	.03	.12
Multi-media	MM	.03	.03	.04	.02	.02
Assessment	A	.11	.17	.08	.07	.13
Pedagogical Moves						
Humor	HUM	.10	.17	.11	.07	.06
Anecdote/example	ANEX	.22	.18	.29	.24	.18
Graphic	GR	.52	.63	.63	.47	.37
Organization	ORG	.10	.12	.07	.09	.10
Emphasis	EMP	.05	.09	.03	.03	.04
Teacher-Student Interactions						
Instructor rhetorical questions	IRQ	.12	.10	.10	.12	.17
Instructor display questions[b]	IDQ	.36	.46	.39	.31	.29
Comprehension questions	CQ	.07	.09	.07	.05	.06
Student novel question	SNQ	.04	.03	.02	.08	.02
Student comprehension question	SCQ	.11	.06	.20	.12	.09
Student response[c]	SR	.28	.36	.29	.25	.21
Student peer interactions	PI	.11	.10	.10	.11	.14
Cognitive Engagement						
Problem-solving	PS	.15	.11	.17	.17	.19
Creating	CR	.02	.03	.01	.04	0
Connecting to real-world	CN	.25	.21	.30	.31	.20
Instructional Technology						
Chalkboard	CB	.19	.01	.32	.15	.39
Overhead projector	OP	.08	.04	.19	.05	.07
PowerPoint	PP	.57	.86	.25	.56	.41
Clickers	CL	.10	.16	.08	.04	.14
Demonstrations	D	.02	.01	.05	.01	.04
Digital tablet	DT	.10	.11	.21	.03	.08
Movies	M	.03	.03	.03	.02	.02
Simulations	SI	.01	0	.02	0	.01

[a] i.e., 2+ questions posed, [b] i.e., seeking new information, [c] i.e., to teacher question.

Frequency of individual codes. The data in Table 5 represent the proportion of times that each TDOP code was observed across all two-minute intervals for the entire sample as well as for each disciplinary group.

Table 6. **Classroom observation data using the TDOP by discipline and course level**

	Code	Course Level		Class Size			
		Upper	Lower	<25	26-100	100-199	200+
Instructors							
Intervals		1030	1484	344	828	821	521
Teaching Methods							
Lecturing	L	.07	.06	.10	.06	.08	.02
Lecturing w/premade visual	LPV	.60	.68	.48	.71	.55	.82
Lecturing w/handmade visual	LHV	.33	.24	.36	.18	.42	.15
Lecturing w/demonstration	LDEM	.04	.03	.04	.00	.07	.03
Socratic lecture[a]	SOC-L	.05	.02	.02	.14	.07	.02
Working through problems	WP	.06	.09	.03	.09	.08	.06
Small group work	SGW	.15	.09	.13	.12	.09	.13
Desk work	DW	.08	.06	.15	.02	.08	.05
Multi-media	MM	.04	.02	.02	.03	.03	.01
Assessment	A	.10	.13	.06	.05	.15	.19
Pedagogical Moves							
Humor	HUM	.10	.11	.06	.06	.13	.18
Anecdote/example	ANEX	.21	.23	.27	.25	.21	.18
Graphic	GR	.49	.53	.35	.46	.57	.54
Organization	ORG	.08	.11	.09	.10	.08	.12
Emphasis	EMP	.03	.06	.03	.04	.04	.10
Teacher-Student Interactions							
Instructor rhetorical questions	IRQ	.07	.15	.07	.14	.11	.12
Instructor display questions[b]	IDQ	.35	.37	.34	.27	.40	.47
Comprehension questions	CQ	.06	.07	.03	.09	.06	.03
Student novel question	SNQ	.02	.05	.02	.04	.06	.03
Student comprehension question	SCQ	.13	.10	.11	.14	.12	.06
Student response[c]	SR	.28	.29	.31	.21	.30	.37
Student peer interactions	PI	.14	.09	.11	.10	.09	.15
Cognitive Engagement							
Problem-solving	PS	.17	.14	.17	.14	.16	.17
Creating	CR	.05	.01	.00	.06	.00	.00
Connecting to real-world	CN	.24	.28	.30	.30	.24	.21
Instructional Technology							
Chalkboard	CB	.25	.14	.36	.17	.25	.00
Overhead projector	OP	.07	.09	.08	.05	.13	.04
PowerPoint	PP	.47	.63	.40	.60	.44	.81
Clickers	CL	.08	.12	.00	.05	.14	.19
Demonstrations	D	.03	.02	.04	.00	.05	.01
Digital tablet	DT	.12	.09	.00	.06	.12	.19
Movies	M	.04	.02	.00	.04	.03	.02
Simulations	SI	.00	.01	.06	.00	.02	.00

[a] i.e., 2+ questions posed, [b] i.e., seeking new information, [c] i.e., to teacher questions

These data highlight the prevalence of certain instructional practices across the five TDOP dimensions. Notable results include an extensive amount of lecturing with pre-made visuals (64% of all two-minute intervals), lecturing with hand-made visuals (27%), the administration of assessments (11%), and small group work (11%). Other important aspects of teaching are the pedagogical strategies not tied to any particular teaching method including the use of anecdotes and examples (22%) and organizational markers (10%). Similarly, faculty employ different approaches to interacting with students in the classroom, particularly through the posing of different types of questions including rhetorical questions where students are not expected to answer (12%) or more open-ended questions, known as display questions that solicit specific information from students (36%). Different types of student cognitive engagement were also documented (e.g., problem solving at 15%), as well as different instructional technologies used in the classroom. Variations among the disciplinary groups included in the study sample are also evident for teaching behaviors such as different types of lecturing, question-posing, and the use of anecdotes. Given that teaching practices may also vary according to other factors such as course level and class size, Table 6 includes data grouped according to these variables.

As with disciplinary groupings, the results indicate differences in teaching practice such as the use of small group work in lower-division courses (9%) and upper-division courses (15%), and the posing of display questions among different class sizes. Future research in this area should explore in greater depth factors that are associated with differences in teaching practice such as those mentioned here as well as gender and appointment type.

Social network analysis graphs. While these data provide a fine-grained account of classroom practice, in de-composing teaching into single, isolated variables, the fiction that teaching can be adequately represented by such measures is perpetuated. Instead, a more accurate approach that aligns with systems-of-practice theory is to identify how combinations of codes were observed at the same time. To address this issue I used social network analysis techniques to depict two instructors data (see Figure 1).

The graphs depicted here illustrate two key points. First, the fact that individual codes do not exist in a vacuum but instead often co-occur within the same two-minute interval is clearly illustrated by these graphs. The lines connecting the codes vary in thickness (on a scale of 1-5) depending on the number of times each pair of codes was observed in the same interval. As a result, the thicker lines indicate an increased frequency with which codes were observed during the same interval. The codes are also positioned in ways that minimize the variation in line length, such that codes closer to the center of the graph tend to be those that are more frequently observed together. Second, the graphs highlight the limitations with simplistic terms such as “lecturing,” because both faculty exhibited a high degree of lecturing with pre-made visuals (80% and 78% respectively), yet their overall teaching methods vary considerably. Instructor #1 relied on lecturing with PowerPoint slides as her primary mode of instruction, with other behaviors (e.g., using anecdotes) more peripheral and thus infrequently used in her class. Instructor #2 also relied on lecturing with PowerPoint slides but utilized a far more diversified range of teaching methods, pedagogical strategies, and student-teacher interactions. Thus, the graphs demonstrate that while both faculty lectured for extensive periods, their teaching approach and thus the learning environment for their students were quite different.

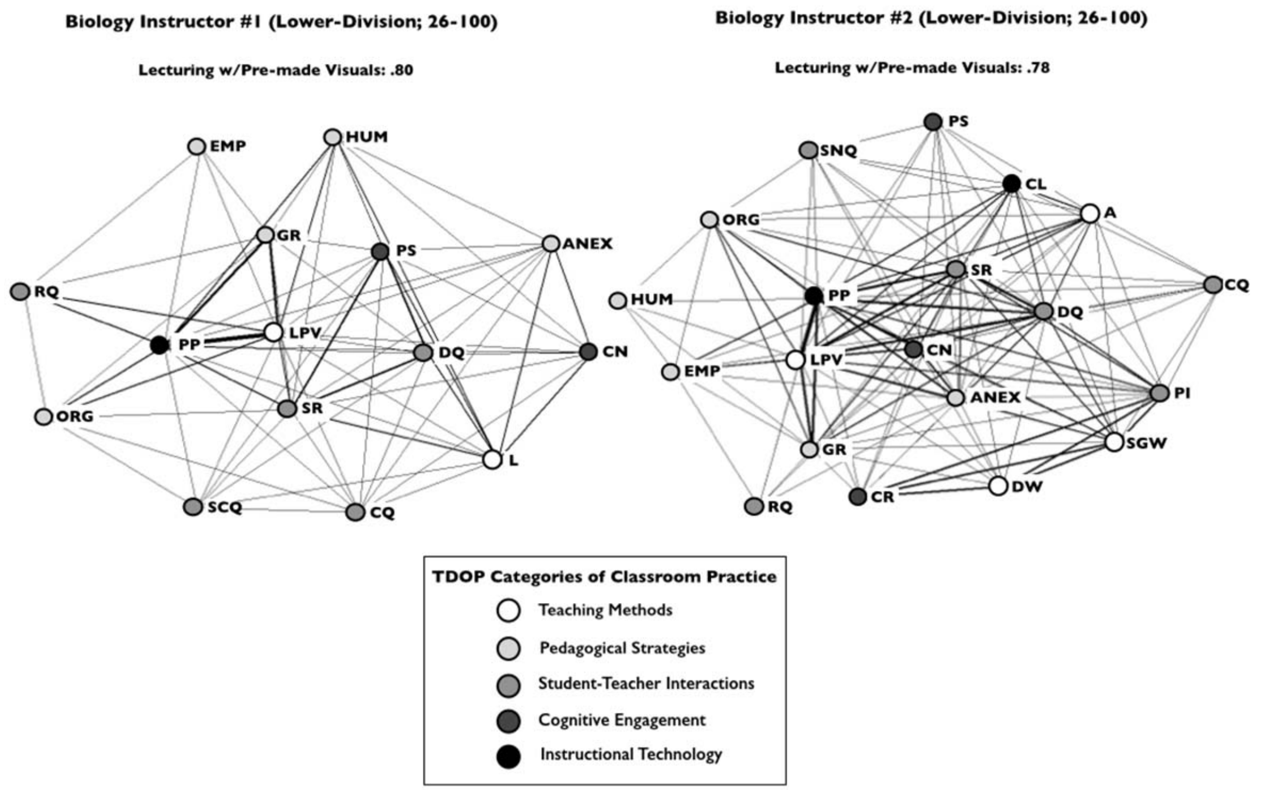


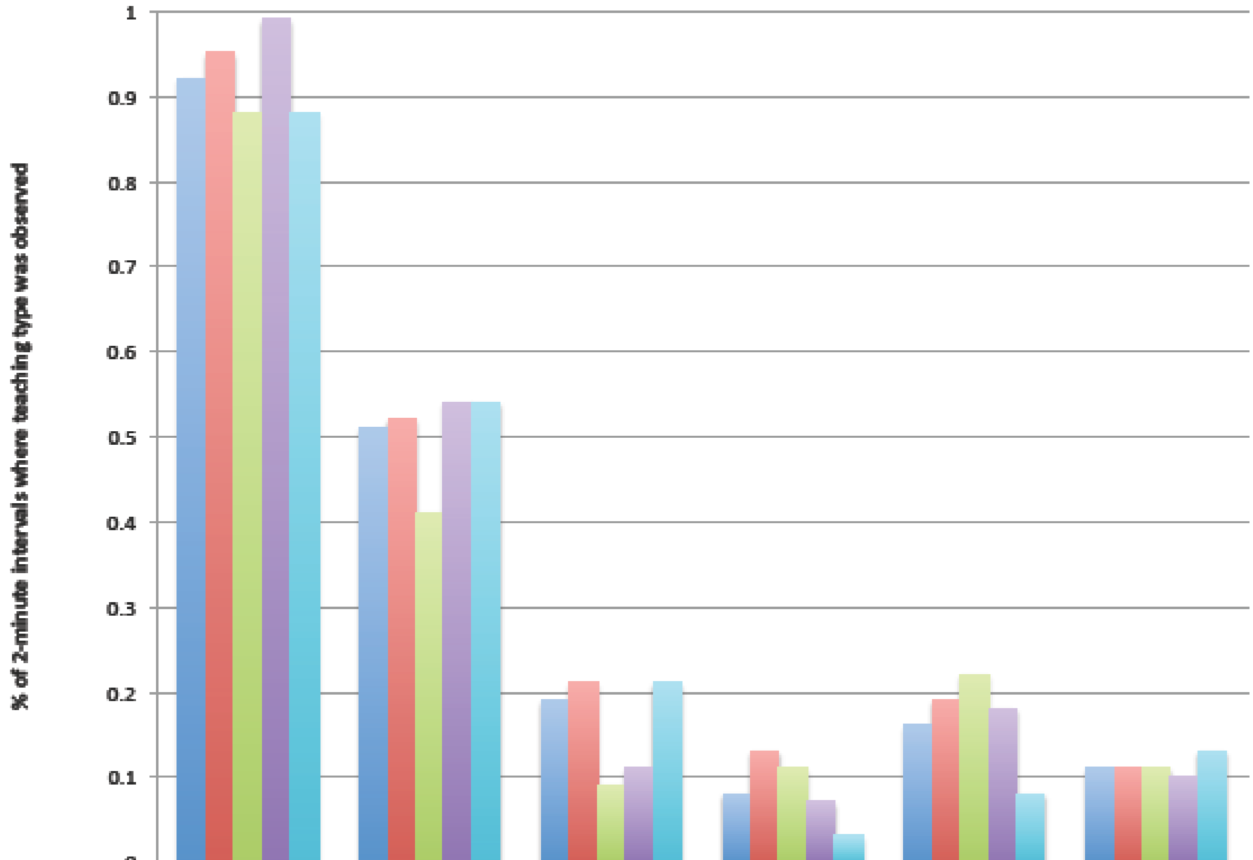
***Figure 1*. Social network analysis graph for two instructors**

2. To what degree are faculty using interactive teaching practices?

Next, I report data that address the primary question addressed in this paper—how can TDOP data indicate the prevalence (or lack thereof) of interactive teaching methods?

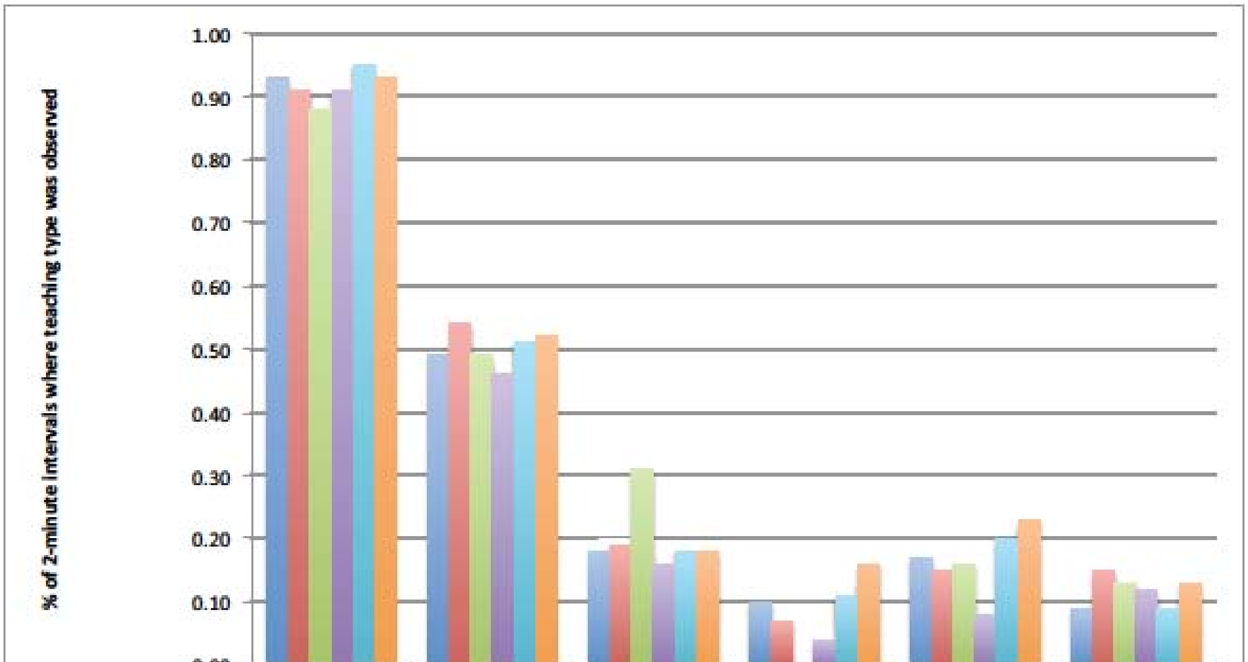
Teacher- and student-centered instruction. One of the ways postsecondary researchers have characterized teaching is whether an instructor adopts a teacher- or student-centered approach (Kember, 1997). Empirical research using such a characterization tends to be survey-based, and few examples exist for how to operationalize observation data to capture these broad categories. With the full acknowledgement that collapsing multiple codes to describe teaching in this manner is a coarse and simplifying exercise, given that more subtle features of instruction are obscured, I report findings using these constructs in Figures 2 and 3.

The results indicate a relatively high degree of teacher-centered instruction across disciplines, course levels, and class sizes. The results also indicate that students are engaged in student-centered activity for approximately half of the observed class periods. It is important to recognize that because a teacher-centered practice (e.g., LHV) and a student-centered practice (e.g., SGW) could be observed within the same two-minute interval, these categories do not sum to 100.



	Teacher-Centered	Student-Centered	Q&A (Verbal)	Q&A (Elec)	Student Activity (Solo)	Student Activity (Group)
Total	0.92	0.51	0.19	0.08	0.16	0.11
Biology	0.95	0.52	0.21	0.13	0.19	0.11
Physics	0.88	0.41	0.09	0.11	0.22	0.11
Geoscience	0.99	0.54	0.11	0.07	0.18	0.1
Mech. Eng.	0.88	0.54	0.21	0.03	0.08	0.13

***Figure 2*. Indicators of interactive teaching by disciplinary group**



	Teacher-Centered	Student-Centered	Q&A (Verbal)	Q&A (Elec)	Student Activity (Solo)	Student Activity (Group)
Course Level Lower	0.93	0.49	0.18	0.10	0.17	0.09
Course Level Upper	0.91	0.54	0.19	0.07	0.15	0.15
Class Size <25	0.88	0.49	0.31	0.00	0.16	0.13
Class Size 26-100	0.91	0.46	0.16	0.04	0.08	0.12
Class Size 101-199	0.95	0.51	0.18	0.11	0.20	0.09
Class Size 200+	0.93	0.52	0.18	0.16	0.23	0.13

***Figure 3*. Indicators of interactive teaching by course level and class size**

More fine-grained indicators of interactive teaching. When interactive teaching is measured at a finer level of granularity where the TDOP codes more closely approximate discrete teaching strategies, the results are unsurprisingly different. In each case, levels of question-and-answer exchanges and student task-based activities are lower than the more global measure of student-centered instruction described above.

3. How, if at all, can time-sampled data be used to provide insights about teaching?

As previously mentioned, one of the strengths of the TDOP instrument is that it collects temporally sensitive data, which is important given that over the course of a class period an instructor will use different teaching practices and these may be used in different combinations with one another. To illustrate the temporal flow of instruction, one class period was visualized where selected indicators of interactive teaching (i.e., teacher- and student-centered instruction, and question-and-answer exchanges) as well as other facets of effective instruction (e.g., organization) are depicted for each two-minute interval (see Figure 4). In addition, observer notes are provided for major class activities.

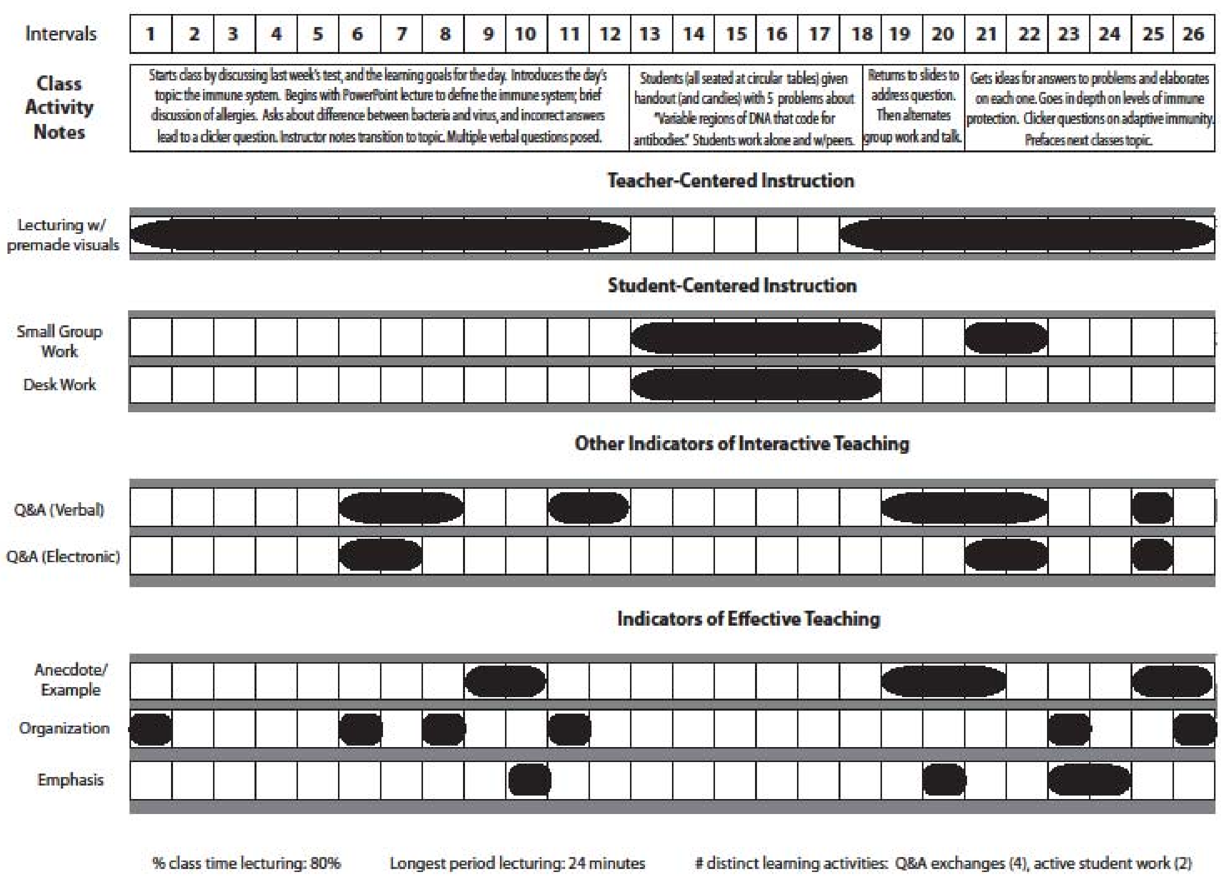
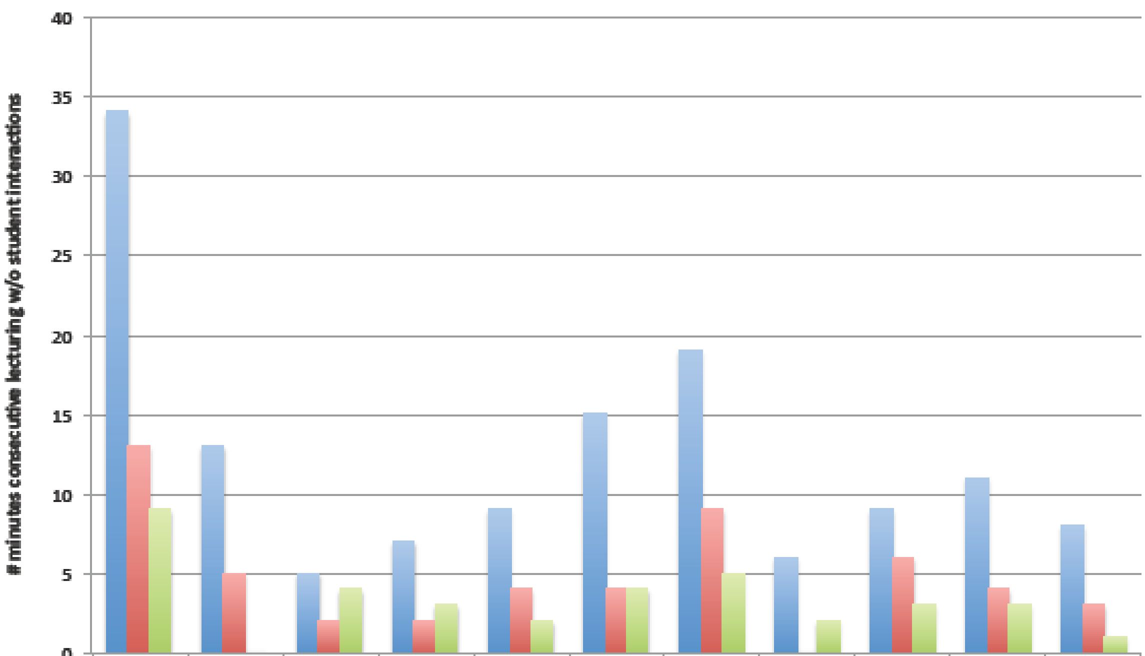


Figure 4. **TDOP data depicted in a time-series graph for two instructors**

One of the key findings from this graphic is that the instructor used a lecturing modality extensively (i.e., in 78% of the two-minute intervals), but the class was centered around an interactive activity where students worked in groups and alone on a problem-based exercise. In addition, the verbal exposition part of the class included numerous questions (both verbal and electronic) as well as indicators of effective teaching, further underscoring how a top-line measure of "lecturing" obscures other aspects of teaching.

Then, given the interest in utilizing the lecture method for relatively short periods (e.g., Handelsman, Miller & Pfund, 2007), the TDOP data were analyzed to categorize instructors based on the length of time that they used a lecturing modality without any form of student engagement (e.g., question-asking) (See Figure 5).



	Total	Biology	Physics	Mech Eng	Geology	Upper	Lower	<25	26-100	100-199	200+
		Discipline				Course Level		Class Size			
0-20	34	13	5	7	9	15	19	6	9	11	8
21-40	13	5	2	2	4	4	9	0	6	4	3
41 +	9	0	4	3	2	4	5	2	3	3	1

***Figure 5*. Periods of extended lecturing with no student interactions by group**

The results indicate that 61% of the sample (n=34) lectured with no student interactions for periods of 20 minutes or less, 23% (n=13) lectured for periods between 21 and 40 minutes, and 16% (n=9) lectured for over 40 minutes.

With insights gleaned from these temporal analyses, I suggest that the TDOP data can be used for a variety of purposes including the identification of underlying characteristics of a class (e.g., dosage, sequence) to carefully assign experimental conditions and to provide faculty with rigorous data for reflective practice.

Discussion

The overall goal of this study was to advance a new approach for studying instructional practice in higher education, while also providing new insights into current teaching practices for a sample of science and engineering faculty. In this section I discuss key findings regarding faculty teaching practices and ways in which the TDOP and its analytic techniques may contribute to the field of science education.

Considering the Classroom Practices of 56 Science and Engineering Faculty

While the results reported in this paper reflect the teaching of a small group of faculty, the results do provide insights into the nature of science and engineering teaching in general. In order to situate the results within the broader literature I compare the data with prior research, though direct comparisons are not possible given the different research instruments utilized and differences in study populations.

Lecturing remains a deeply entrenched instructional modality. The results indicate that faculty in the study sample utilized different forms of lecturing with some regularity, particularly in conjunction with pre-made visuals (observed in 64% of all two-minute intervals). In terms of extensive periods of lecturing with no visible student engagement, 13 faculty (23% of the total sample) were observed lecturing for periods between 21 and 40 minutes, and nine (16%) lectured for over 41 minutes with no student interactions. Based on these findings, it is clear that the lecture is alive and well in many postsecondary classrooms. That said, because 34 faculty (61% of the total sample) lectured for shorter periods (less than 20 minutes), the data indicates that a majority of faculty in the study sample were not engaged in extensive lecturing that lacked any interactions with students. Further, some instructors also interspersed questions, small-group work, and other activities throughout the class, with lecture being used to present content, set up group activities, and provide an organizational scheme for the lesson. Such a pedagogically informed use of lecturing was evident in the case of the biology instructor featured in the time-sampling graph and is the type of practice described by Saroyan and Snell (1997) when they made the following observation:

> A lecture can be as effective as any other instructional strategy so long as it is appropriately suited to the intended learning outcomes and is pedagogically planned and delivered (p. 102).

As a result, while the data indicate a substantial use of the lecturing modality, it cannot be immediately assumed that this equates with inferior instruction and poor student learning[5]. Indeed, in practice Scientific Teaching does not entirely preclude the use of lecturing, but instead aims for only 34% of the class period devoted to lecture, all of which should ideally be broken

[5] While some may suggest that definitive experimental evidence exists supporting this point (e.g., Freeman et al, 2014), I argue that the lack of operationally precise definitions for what constitutes the lecturing condition in many studies renders such judgments premature. The inability to discern precisely what instructional behaviors were used in control and experimental conditions in studies such as Deslauriers, Schelw and Wieman (2011) raise questions about what, in fact, is being compared at all, as well as the validity of the results themselves. No amount of statistical modeling (e.g., random effects models used for meta-analyses that address variable implementation of a treatment) can ameliorate this fundamental design problem.

into segments no longer than 10 minutes (Miller et al, 2008). In Peer Instruction classrooms, lecturing also takes place, as Crouch and Mazur (2001, p. 975) note, “We typically devote one-third to one-half of class time to ConcepTests and spend the remainder lecturing.” In this case, the lecture serves as a way to introduce new topics and prepare students for more in-depth activities. Thus, the pertinent question becomes one of the amount and timing of lecturing, as well as whether an underlying pedagogical rationale is guiding its use. While the specific duration and type of lecturing that is the most pedagogically beneficial remains an open question for the field to pursue in the future, the weight of the evidence regarding the importance of active engagement does suggest that there is room for greater incorporation of non-lecture activities in some postsecondary classrooms, particularly in the case of those faculty who continually lectured for over 41 minutes with nary a question or activity.

In regard to how the data compare with the extant literature on the prevalence of lecturing, such direct comparisons are not useful given the fact that it is impossible to determine what is meant by lecturing in many surveys and whether respondents understand the term in the same manner (Porter, 2011). However, some tentative conclusions may be drawn. While the HERI Faculty Survey (Hurtado et al, 2012) does not define what is meant by the term “extensive lecturing,” the 63% of respondents who indicated this option as a regularly used teaching method is similar to the results for the use of pre-made visuals using the TDOP instrument (observed in 64% of all two-minute intervals), but varies from results capturing long periods of lecturing. This suggests the possibility that survey respondents may interpret the phrase “extensive lecturing” not in terms of consecutive minutes of use but regarding its overall prevalence within a class. In any case, the problems inherent in such comparisons highlight the fact that despite calling for the field to operationally define “lecturing” over 20 years ago, Schonwetter’s (1993) call has not yet been answered. Considering the challenges in the STEM education literature with the careful operationalization of lecturing as a treatment used in experimental designs, it is clear that the field should address this problem with the same care and precision that scientists utilize in their own disciplinary research (Derting et al, 2011).

The extent of interactive teaching practices. The challenges associated with survey response processes in regard to lecturing also apply to interactive teaching methods. That is, do survey respondents understand terms such as “active learning” or techniques such as Peer Instruction in the same manner? Challenges with measuring interactive teaching, however, are not limited to surveys but also apply to observation research. As previously noted, data on interactive teaching obtained using the TDOP are necessarily limited in two ways: (1) they do not indicate the quality with which interactive methods were used in the classroom, and, (2) they capture only a small portion of what pedagogical behaviors (e.g., course planning, classroom instruction, laboratory sessions) may actually comprise an interactive teaching approach. With these limitations in mind, however, the data reported in this paper do provide some empirically based insights into the extent of certain aspects of these modalities.

Using a blunt measure of interactive teaching (i.e., teacher- or student-centered instruction) gives the analyst the equivalent of a 30,000-foot perspective on classroom practice. The data indicate that teacher-centered practices were extensively observed (in 92% of all two-minute intervals) and student-centered practices were observed in roughly half of the courses (51%). Again, because within any interval both types of teaching can occur and thus are not mutually exclusive categories, the data indicate what while in most cases the teacher is maintaining “the floor” in terms of control of the classroom discourse, a considerable amount of student-led activity is taking place within such a context. For more specific indicators of interactive teaching, the data indicate that verbal questions were observed in 19% of all two-minute intervals, clicker questions in 8%, student solitary work (i.e., desk work) in 16% and student collaborative or group work in 11%. With these results, it appears that certain elements of the practices advocated by policymakers and educators are in fact being used in the field. These results suggest that students may be spending an appreciable amount of time in interactive teaching modalities than often assumed, and that is conveyed by headlines in the media such as “Lectures still dominate math and science teaching” (Berrett, 2012).

In comparing these data with the literature, keeping the caveats regarding response process issues and comparability in mind, the results are both consistent and contradictory. First, the HERI survey showed that 47% of faculty used small groups and 61.5% used class discussions (Hurtado et al, 2012), whereas the results reported here indicate smaller use of these methods (11% and 0% respectively). In the Henderson and Dancy (2009) survey, 13.9% of the respondents reported using Peer Instruction and 13.7% reported using group work. While no TDOP codes reflect Peer Instruction in strict terms, the use of clickers (8%) and small group work (11%) capture key facets of the strategy and similar to findings in other studies. Similarly, Prince et al (2013) found Peer Instruction to be used by 15% of the respondents, which was relatively close to the current results, though the current findings that problem based learning was used by 35% and active learning by 60% were dissimilar to their survey results.

The TDOP Instrument: Contributions to the Field of Science Education

As previously noted, since the TDOP was first used in a research study and results disseminated, a persistent question has been, “So what?”—particularly in regard to how it could provide insights into the prevalence of interactive teaching. In response, despite considerable limitations with the TDOP as well as other observation protocols, I highlight the various uses to which educators can put fine-grained data that mirrors the complexity of real-world classrooms, particularly in the support of faculty development efforts.

Perhaps the most significant contribution of the TDOP is its ability to capture multiple aspects of instruction at a fine-grained level as they interact and unfold over time. Such an instrument is consistent with theories of educational practices (Spillane, Halverson & Diamond, 2001) that reject the view that educational practice is the sole domain of a single actor regardless of the social or organizational context. In breaking down instruction into constituent parts, the TDOP builds on prior work exploring nuanced aspects of faculty teaching (e.g., Murray, 1983), and ideally work on further refining this approach to capture specific pedagogical techniques and instructional quality will continue.

In addition to capturing the prevalence (or lack thereof) of interactive modalities, the TDOP and related analytic techniques used in this paper contribute to the field a way to assess current faculty practices, as well as their sequencing with other instructional modalities and their intensity of use. In particular, the time-sampling approach to studying postsecondary classrooms pioneered by the TDOP allows users to capture aspects of sequencing and dosage in a structured manner that previously were unavailable to higher education researchers. While the time-sampling framework itself is critical, I argue that without nuanced codes that extend beyond student-and teacher-centered instructional methods (e.g., Smith et al, 2013), the resulting data reflect a relatively blunt measure of teaching that may obscure more that it reveals.

Finally, one of the most promising uses to which TDOP data can be put is that of supporting and enhancing reflective practice as part of faculty development efforts. One of the critical features of effective professional development activities for postsecondary faculty is the provision of credible, detailed and timely feedback for instructors that can spark critical reflection on their own teaching (Chism, 2007; Gormally, Evans & Brickman, 2014). Yet data about teaching, beyond the ubiquitous end-of-term student evaluations, are often in short supply in many colleges and universities, and so faculty are often left with little or no data upon which to consider whether their teaching was effective or not. Such critical reflection is a cornerstone to the ongoing development of professional expertise (Schön, 1983), and I suggest that the TDOP can be particularly useful in providing data for these purposes. One of the benefits of the TDOP is that being a descriptive instrument faculty may be more responsive and less threatened to the results in contrast to evaluative protocols such as the RTOP. In regard to the types of data faculty find most useful for these purposes, in an unpublished field test of the utility of TDOP data to promote self-reflection, a biology teacher found that time-series graphs (e.g., Figure 4) were the most useful in shedding light on the sequence of learning activities. Building on the prospects for these data to support faculty professional growth and development, my research group is currently in field-testing whether a combination of TDOP and student evaluation data can in fact be useful for these purposes.

Limitations. Several limitations to the TDOP (and thus the data reported in this study) should be acknowledged. First, being a descriptive instrument the TDOP does not shed any light on whether or not any particular teaching behaviors are being used to good effect. Additional sources of data regarding the efficacy of certain instructional practices (e.g., assessments of learning gains) are needed to complement the insights provided by the TDOP in order to arrive at any estimation of instructional quality. Second, one of the major limitations with observation-based data is that it relies on the observer to infer whether or not a particular behavior has occurred. While this is less problematic in regards to capturing discrete, easily interpreted phenomenon (e.g., the use of instructional technology) it becomes a significant issue if the intent of an observation is to estimate phenomena such as potential student cognitive engagement. Thus, there remains a certain degree of error associated with any given code frequency. Third, while the training described in this paper was rather extensive, only 2% of the dataset were used to establish IRR. In the future, training should include more videotaped lectures to test IRR (e.g.,10 -15) though trade-offs with the increased time for training should be considered. Finally, limitations to the study reported in this paper include the self-selected nature of the sample, the lack of observations conducted throughout the course of a term, and the lack of data on laboratory or discussion sections.

Implications for Policy and Educational Reform

In studying classroom teaching, the field of undergraduate science education is shifting from a reliance on self-reported data obtained via questionnaires to classroom observations (Ebert-May et al, 2011; Smith et al, 2013). This methodological shift is important because a more rigorous and nuanced accounting of instructional practice is warranted as increasing amounts of resources are allocated to "transforming" how undergraduate courses across all disciplines are taught. Without evidence about the nature of current teaching and whether or not it is changing over time, the ultimate efficacy of these investments remains unknown. But perhaps most importantly, such first-hand evidence on teaching drives home the point that teaching must be documented and described in complex terms, which should pave the way for the field to begin having more nuanced and empirically-based conversations about faculty teaching practices.

Such conversations should be grounded in the simple fact that for many scholars of postsecondary teaching, lecturing as an all-encompassing descriptor of classroom teaching is neither credible nor evidence-based (Schonwetter, 1993; Saroyan & Snell, 1997). Instead, as I demonstrate in this paper, verbal exposition varies along multiple dimensions including dosage, its sequencing with other instructional strategies, and co-occurrence with additional dimensions of instructional practice such that two faculty who "extensively lecture" can craft radically different learning experiences for their students (see Figure 1).

Yet the current rhetoric used by many policymakers and researchers active in STEM education reform perpetuates the mistaken notion that the term "lecturing" refers to a distinct type of instructional practice, instead of one that varies in a multitude of ways. Further, this reductionist view of verbal exposition is used to create the binary categorization that pits lecturing against interactive teaching, with the former indicative of poor instruction and the latter of high-quality instruction (PCAST, 2012; Freeman et al, 2014). Some have taken the value judgments ascribed to this crude categorization scheme further in labeling the lecture method to be the "pedagogical equivalent of bloodletting" and its use unequivocally leading to an "inferior education" (Wieman, 2014, p. 8320).

In creating a simple dichotomy into which faculty can be placed, the current rhetoric used in STEM education is not dissimilar from the ways that previous scholars have conceptualized faculty cognition and instructional approaches as either teacher-centered or student-centered (e.g., Kember, 1997). However, this long-standing dichotomy has come under significant challenge in the past decade, with some arguing that this view ignores different stages of cognition relative to the context and task situations (McAlpine et al., 2006), and that both faculty thinking and classroom teaching is sufficiently complex that "a strong opposite 'either/or' positioning of the approaches does not do justice to the nature of the phenomenon" (Postareff & Lindblom-Ylanne, 2008, p.120). Decades of research on teacher cognition in K-12 schools support this position (e.g., Borko, Roberts & Shavelson, 2008).

Why is a more nuanced conversation about teaching important? Because conceptions about the nature of faculty teaching based on ill-defined binary categorizations are being used to inform both educational policy and practices in faculty development (Devlin, 2006). Based on the evidence reported in this paper I argue that the field of science education stands to benefit from a more careful discussion of undergraduate teaching in at least two ways.

First, in some educational experiments the treatment being utilized is often an ill-defined and ambiguous form of "lecturing," such that the precise nature of the experimental conditions is uncertain (e.g., Deslauriers et al, 2011). Since evidence from studies such as this are being cited in policy documents (e.g., PCAST, 2012), it is incumbent upon the field of science education to design empirical studies with the same care and attention to rigor that disciplinary scientists utilize in their own basic research. With tools such as the TDOP it becomes possible for researchers to more carefully delineate precise types of instructional situations, which will be necessary if the field wishes to move beyond the lecturing and interactive teaching dichotomy to what some are calling "second-generation research" (Freeman et al, 2014).

Second, research on reform implementation indicates that recipients of an intervention are most receptive when the messaging accompanying the innovation, not to mention the innovation itself, are closely aligned with the existing practices, cultural traditions, and beliefs of the population (Rogers, 2010; Spillane, Reiser & Reimer, 2002). There does exist some evidence that STEM faculty are having a negative response to the approach taken by some educational reformers, who are seen as adopting an overly top-down approach that ignores local practices and implicitly views the faculty member as the primary problem (Henderson & Dancy, 2008). Ultimately, I speculate that reforms that promote interactive teaching may be more readily adopted, or at least seriously considered, if the messaging used by advocates reflects an understanding of the actual teaching practices used by faculty in their daily work—which the use of the lecturing versus interactive teaching framing fails to do. Indeed, given that a majority of faculty in the study reported in this paper relied on some form of verbal exposition in their classes, it is possible that suggesting slight modifications to the lecturing method may be a more promising approach than calling for the outright transformation of an instructor's entire pedagogical approach (Martin & Ramsden, 2003). Thus, it may not be a matter of eliminating lecture from one's pedagogical toolkit, but instead the problem is how to alter one's lecturing approach to have more of a deliberate pedagogical purpose while also not boring one's students to tears. An extensive amount of research is underway on this point (e.g., Walker, Cotner, Baepler, & Decker, 2008), and I suggest that the field will be better served in making minor yet influential adaptations to how verbal exposition is used in the classroom, rather than advocating for its complete and utter elimination.

In making these points, I am not questioning the value of interactive instruction or advocating for the use of extensive lecturing. Instead, I am claiming that teaching is too complicated to be boiled down to a single term and that to reach the goal of providing a high-quality learning environment for all students, a more nuanced approach to thinking about instruction is a necessary precursor to communicating with faculty and advancing our understanding of the effects of different pedagogical approaches.

References

Berrett, D. (2012, October 25). Lectures still dominate science and math teaching, sometimes hampering student success. *The Chronicle of Higher Education*.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *UCINET for Windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Borko, H., Roberts, S. A., & Shavelson, R. (2008). Teachers' decision making: From Alan J. Bishop to today. In *Critical issues in mathematics education* (pp. 37-67). Springer US.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, and school.* Washington, DC: National Research Council.

Brown, G. & Bakhtar, M. (1988). Styles of lecturing: A study of its implications. *Research Papers in Education 3*(2), 131-153.

Cash, A. H., Hamre, B. K., Pianta, R. C., & Meyers, S. S. (2012). Rater calibration when observational assessment occurs at large scales: Degree of calibration and characteristics of raters associated with calibration. Early Childhood Research Quarterly, 27(3), 529–542.

Chism, N.V.N. (2007). *Peer review of teaching: A sourcebook* (2nd Ed.). Anker Publishing Company: Bolton, MA.

Cole, M. (1996). *Cultural psychology: A once and future discipline*. Harvard University Press: Cambridge, MA.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics, 69,* 970-977.

Danielson, C. (2013). *The framework for teaching evaluation instrument* (2013 Edition). The Danielson Group.

Derting, T., Williams, K. S., Momsen, J. L., & Henkel, T. P. (2011). Education research: set a high bar. *Science, 333,* 1220.

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, *332*(6031), 862-864.

Devlin, M. (2006). Challenging accepted wisdom about the place of conceptions of teaching in university teaching improvement. *International Journal of Teaching and Learning in Higher Education, 18*(2), 112-119.

Ebert-May, D., Derting, T. L., Hodder, J., Momsen, J. L., Long, T. M., Jardeleza, S. E. (2011). What we say is not what we do: effective evaluation of faculty professional development programs. *BioScience*, *61*(7), 550-558.

Feldman, K. A. (1989). Association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education, 30,* 583–645.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. (2004). School engagement: Potential of the concept: State of the evidence. *Review of Educational Research, 74,* 59–119.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111* (23) 8410-8415.

Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. *CBE-Life Sciences Education*, *10*(2), 175-186.

Froyd, J. E., Borrego, M., Cutler, S., Henderson, C., Prince, M. J. (2013). Estimates of use of research-based instructional strategies in core electrical or computer engineering courses. *IEEE Transactions on Education, 56*(4), 393-399.

Gormally, C., Evans, M., & Brickman, P. (2014). Feedback about Teaching in Higher Ed: Neglected Opportunities to Promote Change. *CBE-Life Sciences Education, 13*(2), 187-199.

Grunspan, D. Z., Wiggins, B.L., & Goodreau, S.M. (2014). Understanding classrooms through social network analysis: A primer for social network analysis in education research. *CBE-Life Sciences Education, 13*(2), 167-178.

Guarino, C. & Tracy, B. (2012). Review of gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. National Educational Policy Center. Boulder, CO.

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

Handelsman, J., Miller, S., & Pfund, C. (2007). *Scientific teaching*. New York, NY: W.H. Freeman and Co.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., ... & Wood, W. B. (2004). Scientific teaching. *Science, 304*(5670), 521-522.

Hativa, N., & Goodyear, P. (Eds.) (2001). *Teacher thinking, beliefs, and knowledge in higher education.* Norwell, MA: Kluwer Academic Publishers.

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: an analytic review of the literature. *Journal of Research in Science Teaching, 48*(8), 952-984.

Henderson, C. R. & Dancy, M. H. (2009). Impact of physics education research on the teaching of introductory quantitative physics in the United States. *Physical Review Special Topics–Physics Education Research, 5*, 020107.

Henderson, C., & Dancy, M. H. (2008). Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations. *American Journal of Physics, 76*(1), 79-91.

Hmelo-Silver, C. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235-266.

Hora, M. T., & Ferrare, J. J. (2013). Instructional systems of practice: A multidimensional analysis of math and science undergraduate course planning and classroom teaching. *Journal of the Learning Sciences, 22*(2), 212-257.

Hurtado, S., Eagan, K., Pryor, J.H., Whang, H. & Tran, S. (2012). Undergraduate teaching faculty: The 2010-2011 HERI Faculty Survey. Los Angeles, CA: Higher Education Research Institute, UCLA.

Kane, R., Sandretto, S., & Heath, C. (2002). Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research, 72*(2), 177–228.

Kane, M.T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38* (4), 319-342.

Kember, D. (1997). A reconceptualisation of the research into university academics' conceptions of teaching. *Learning and Instruction, 7*, 255–275.

Lee, O., & Brophy, J. (1996). Motivational patterns observed in sixth-grade science classrooms. *Journal of Research in Science Teaching, 33*, 303–318.

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

Martin, E., & Ramsden, P. (1993). An expanding awareness: How lecturers change their understanding of teaching. *Research and Development in Higher Education, 15,* 148-155.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*(1), 29–45.

Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ: Prentice Hall.

McAlpine, L., Weston, C., Timmermans, J., Berthiaume, D., & Fairbank-Roch, G. (2006). Zones: reconceptualizing teacher thinking in relation to action. *Studies in Higher Education, 31* (5), 601-615.

McKeachie, W. (1994). *Teaching tips* (9th ed.). Lexington, MA: D.C. Heath and Co.

Middendorf, J., & Kalish, A. (1996, January). The "change-up" in lectures. *National Teachers Learning Forum, 5* (2), 1-5.

Miller, S., Pfund, C., Pribbenow, C. M., & Handelsman, J. (2008). Scientific teaching in practice. *Science, 322*(5906), 1329-1330.

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

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.

Perry, R.P., & Smart, J.C. (Eds.). (1997). *Effective teaching in higher education: Research and Practice.* New York: Agathon Press.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119.

Porter, S. R. (2011). Do college student surveys have any validity?. *The Review of Higher Education*, *35*(1), 45-76.

Postareff, L., & Lindblom-Ylanne, S. (2008). Variation in teachers' description of teaching: Broadening the understanding of teaching in higher education. *Learning and Instruction*, 18, 109-120.

President's Council of Advisors on Science and Technology (2012). *Report to the President. Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering and mathematics.* Executive Office of the President. Washington, DC.

Prince, M., Borrego, M., Henderson, C., Cutler, S., & Froyd, J. (2013). Use of research-based instructional strategies in core chemical engineering courses. *Chemical Engineering Education, 47*(1), 27-37.

Prosser, M. Trigwell, K., & Taylor, P. (1996). A phenomenographic study of academics' conceptions of science learning and teaching. *Learning and Instruction, 4*, 217-231.

Rogers, E.M. (2010). *Diffusion of innovations* (4th Ed.) Simon & Schuster: New York, NY.

Samuelowicz, K., & Bain, J. D. (2001). Revisiting academics' beliefs about teaching and learning. *Higher Education, 41*, 299–325.

Saroyan, A., & Snell, L. S. (1997). Variations in lecturing styles. *Higher Education, 33*(1), 85-104.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.

Schonwetter, D. (1993). Attributes of effective lecturing in the college classroom. *The Canadian Journal of Higher Education 23*(2), 1-18.

Small, A. (2014, May 27). In defense of the lecture: A good lecturer doesn't just deliver facts but models how an expert approaches problems. *The Chronicle of Higher Education*.

Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE-Life Sciences Education*, *12*(4), 618-627.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of educational research, 72*(3), 387-431.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, *30*(3) 23-28.

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), 020101.

Walker, J. D., Cotner, S. H., Baepler, P. M., & Decker, M. D. (2008). A delicate balance: integrating active learning into a large lecture course. *CBE-Life Sciences Education, 7*(4), 361-367.

Wieman, C. E. (2014). Large-scale comparison of science teaching sends clear message. *Proceedings of the National Academy of Science, 111*(23) 8319-8320.