Science
Education

Toward a Descriptive Science of Teaching: How the TDOP Illuminates the Multidimensional Nature of Active Learning in Postsecondary Classrooms

MATTHEW T. HORA
Department of Liberal Arts and Applied Studies and the Wisconsin Center for Education Research, University of Wisconsin-Madison, Madison, WI 53715, USA

Received 22 October 2013; accepted 17 March 2015
DOI 10.1002/sce.21175
Published online 10 June 2015 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: Detailed accounts of teaching can shed light on the nature and prevalence of active learning, yet common approaches reduce teaching to unidimensional descriptors or binary categorizations. In this paper, I use the instructional systems-of-practice framework and the Teaching Dimensions Observation Protocol (TDOP) to advance an approach to thinking about teaching in science classrooms in more multidimensional terms. Using descriptive statistics and social network analysis, I examine the teaching practices employed by a group of science and engineering faculty ($n = 56$). Results indicate the extensive use of lecturing with premade visuals (observed in 65% of all 2-minute intervals comprising a class). However, the majority of instructors ($n = 34$) lectured for periods of 20 minutes or less. Using the Differentiated Overt Learning Activities (Chi & Wylie, 2014) framework to interpret TDOP codes, the data reveal lower rates of active learning modalities including "being active" (students answering questions, 28%; students problem solving (PS),15%), "being constructive" (students asking questions, 4%; students doing creative tasks, 2%), and "being interactive" (students working with peers to do creative tasks, 2%). Results indicate variation across disciplines and course contexts, that active learning is embedded within PowerPoint lectures, and that small group work exercises are not synonymous with constructivist activities. Implications for research, practice, and policy are discussed. © 2015 Wiley Periodicals, Inc. *Sci Ed* **99:**783–818, 2015

Correspondence to: Matthew T. Hora; e-mail: hora@wisc.edu

INTRODUCTION

The goal of this paper is to introduce the Teaching Dimensions Observation Protocol (TDOP) – and the instructional systems-of-practice framework upon which it is based – and describe how it can be used to document classroom dynamics in a nuanced and rigorous manner. This approach is based on the premise that educational practice is best viewed as "distributed in the interactive web of actors, artifacts, and the situation" as opposed to a phenomenon that can be understood through an individual's behavior regardless of the social and organizational context (Spillane, Halverson, & Diamond, 2001, p. 23). Such a perspective is important and necessary because classrooms are complex environments that involve not only instructor behaviors but also other actors and artifacts, which collectively interact over the course of a class period (Cohen & Ball, 1999). Yet commonly used surveys and observation protocols reduce teaching in ways that collapse time and the dynamism of teaching into coarsely-grained and unidimensional descriptors that largely focus on instructor behaviors alone. Such views of teaching have led to the widespread notion that teaching can be adequately described using terms such as "lecturing" or "active learning" (e.g., Freeman et al., 2014). Unfortunately, this approach leads to another, more problematic view of teaching: the conflation of coarse descriptors of teaching (e.g., lecturing) with particular modes of cognitive engagement. The instructional systems-of-practice framework offers a way to describe, in a more rigorous, fine-grained, and theoretically informed fashion, the complexity of teaching "in the wild" of real-world classrooms.

Why is a descriptive approach to the study of teaching that captures its multi-dimensional nature necessary? First, descriptive research is important to any arena of scientific inquiry to better understand the nature of the phenomenon being studied. For example, it almost goes without saying that biological scientists should describe the subjects of their research (e.g., organisms, animal behaviors, or entire ecosystems) in as precise a manner as possible. It is important that researchers studying problems in educational settings approach issues of research design and measurement the same care and precision that scientists utilize in their own disciplinary research (Derting, Williams, Momsen, & Henkel, 2011). Furthermore, given evidence that instructional activities that foster students' active engagement with the material and one another (hereafter referred to as a broad class of activities called "active learning") are more effective than passively listening to a lesson (Bransford, Brown, & Cocking, 1999), researchers are being encouraged to document the degree to which active learning modalities are being used, particularly in college and university classrooms (American Association for the Advancement of Science [AAAS], 2012). Yet currently available research instruments are not designed to describe and document nuances of classroom practices in general and subtle features of active learning in particular at a fine-grained level, particularly as they occur in naturalistic settings.

A second reason why descriptive research on classroom teaching is warranted pertains to the challenges associated with educational reform, particularly at the postsecondary level. While some signs indicate an increase in the adoption of active learning techniques (Hurtado, Eagan, Pryor, Whang, & Tran, 2012), other evidence suggests that instructors are not widely taking up these instructional practices (Henderson & Dancy, 2009; PCAST, 2012). Similar to educational reform efforts in K-12 settings, this state of affairs suggests that a gap exists between what researchers and policymakers consider "best practices" and what is actually being done in the classroom, or what is commonly viewed as the "fidelity" with which these practices are actually being used in the classroom. Yet coarsely-grained approaches to the study of teaching are not sufficient for assessing the "uptake" of active learning techniques because these rather complex approaches to curriculum and instruction focus not only on changing "surface structures and procedures" but also on altering instructors' beliefs, norms of student-teacher interactions, and pedagogical principles embodied in the curriculum (Coburn, 2003, p. 4). Finally, in-depth accounts of teaching can be used as the basis upon which to design new programs or interventions so that they are responsive to local practices instead of top-down mandates which may alienate instructors (Cohen & Ball, 1999; Spillane et al., 2001).

Thus, the field of science education would benefit from a way to describe and document classroom teaching in a way that maintains fidelity to its complex and dynamic nature, while also being able to discern the presence (or absence) of active learning modalities. In response, the TDOP was created as part of a study that explored the cognitive, cultural, and contextual characteristics shaping instruction in postsecondary institutions.[1] The TDOP captures key elements of instructional systems-of-practice theory through its focus on fine-grained descriptors of classroom dynamics, interactions among actors and artifacts, and the temporal fluctuation of these phenomena over time. Since its initial development in 2008, the instrument has been widely adopted with users particularly appreciating the detailed and nonevaluative nature of the TDOP based in large part on the view that postsecondary faculty[2] are especially resistant to having the "quality" of their teaching be determined by a single score or rubric (Chism, 2007) and also the perceived need in the field for more rigorous methods to measure teaching as an empirical phenomena (AAAS, 2012). The TDOP has been featured in several research papers (e.g., Clark, Norman, & Besterfield-Sacre, 2014; Code, Piccolo, Kohler, & MacLean, 2014; Finelli, Daly, & Richardson, 2014), adapted by other researchers (Smith, Jones, Gilbert, & Wieman, 2013)[3] used in numerous studies focused on science education, and over 300 researchers have used the online version of the instrument (http://tdop.wceruw.org).

While research conducted with the TDOP has shed light on the insufficiency of descriptors such as "lecturing" (Hora & Ferrare, 2014) and documented in teaching across disciplinary groups (Hora & Ferrare, 2013), one issue has been underexamined—whether the instructional systems-of-practice framework and the TDOP can contribute insights to the prevalence and nature of active learning. In this paper, I describe techniques for combining TDOP codes to detect the presence of active learning using the Differentiated Overt Learning Activities (DOLA) framework, which is a taxonomy of classroom-based active learning modalities based on observable student behaviors (Chi & Wylie, 2014). Using the DOLA framework to organize combinations of TDOP codes, I analyzed data from observations of 56 instructors using descriptive statistics and social network analysis to answer two research questions: (1) What teaching practices are employed by a group of science and engineering faculty? (2) What is the prevalence and nature of active learning observed in these classrooms? In addition, I also analyze the data within different class sizes and course levels based on the role that the organizational context plays in shaping practice as theorized by the instructional systems-of-practice approach.

[1]The development of the TDOP has been a group endeavor, with Joseph J. Ferrare acting as a co-developer from 2008 to 2013. In addition, a team of four researchers including myself, Jana Bouwma-Gearhart, Amanda Oleson, and Jennifer Collins collected data in the spring of 2013 that are reported in this paper. While this paper is single authored the study reported here would not be possible without the contributions of these valued colleagues.

[2]By faculty I mean all people who hold undergraduate teaching positions—whether full- or part-time, tenured, or untenured—in postsecondary institutions with the exception of graduate teaching assistants.

[3]This effort led to a protocol known as the COPUS, which is a minor adaptation of the TDOP that involved removing certain categories (e.g., student cognitive engagement) and re-naming existing codes.

BACKGROUND

In recent years, the postsecondary science classroom has been the focus of extensive educational reforms promoting instructional activities that more directly engage students in their own learning. Indeed, the federal government views the adoption of active learning in postsecondary science, technology, engineering, and mathematics (STEM) classrooms as a critical national priority associated with economic development, diversity in the workforce, and the public's scientific literacy (e.g., PCAST, 2012). Much of this focus is a reaction to the persistence of didactic lecturing in higher education, where faculty act as a "sage on the stage" by transmitting knowledge to their passive students (Mazur, 2009). This reaction has been fueled by growing evidence in cognitive psychology and the learning sciences (Bransford et al., 1999; Chi & Wylie, 2014), and discipline-based education research (National Research Council, 2012) that instructional activities that directly engage students in actively constructing their own understanding of course material are more effective than passive modes of learning (e.g., Duch, Groh, & Allen, 2001; Crouch & Mazur, 2001).

Given that funding agencies and educational leaders are investing substantial resources to support the nationwide adoption of active learning techniques, these groups have a growing interest in documenting the extent to which the resources are being used in the nation's colleges and universities (AAAS, 2012). This raises a key question: what do we really know about how science is taught in undergraduate classrooms? In the most comprehensive analysis of faculty teaching in the United States, the Higher Education Research Institute (HERI) conducts a survey of faculty work that investigates teaching methods, advising practices, and job satisfaction. In the 2010–2011 survey, which included 23,824 full time undergraduate faculty at 417 postsecondary institutions, researchers found that STEM faculty ($n = 6768$) reporting using "extensive lecturing" in 63% of the courses they taught, "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).[4] Unfortunately, what terms such as "extensive lecturing" means in operationally precise terms is not clear in the survey. A smaller survey-based study focusing on physics instructors found that among 722 instructors, 29% reported using peer instruction, 13.9% reported using interactive lectures, and 13.7% reported using cooperative group problem solving (PS) (Henderson & Dancy, 2009).

While evidence from self-reported surveys such as these shed light on the nature of teaching in STEM classrooms, they are limited in several ways. In a critical review of surveys used to study college student experiences, Porter (2011) highlighted several validity-related issues such as the ability of respondents to accurately and reliably recall their experiences over imprecise periods of time. This concern about the process of respondent recall and the granularity of resulting data is salient for surveys such as the HERI Faculty Survey that asks faculty to report their teaching across "most classes." The widespread reliance on self-reported data to document teaching is also limited by the lack of research exploring the congruence (or lack thereof) between self-reports and actual classroom practice (Kane, Sandretto, & Heath, 2002; Mayer, 1999). Thus, questions remain about precisely what is meant by data such as "63% of faculty are extensive lecturers" in terms of specific classroom behaviors and whether the claim actually reflects respondents' teaching practices.

[4]When the final version of this manuscript was being prepared, data from the 2013–2014 HERI Faculty Survey had not yet been released that disaggregated results by disciplinary group. Thus, in this paper I report data from the 2010–2011 version of the survey.

Problems With Conceptualizing and Operationally Defining Classroom Teaching

But perhaps the biggest issue facing the field pertains to how teaching itself is conceptualized and operationalized in research instruments.

Reliance on Ill-Defined Descriptors of Teaching

One of the challenges in studying teaching is researchers' tendency to characterize instruction solely in terms of teaching methods such as "lecturing" or "small group work." This tendency is problematic because such descriptions frequently lack accompanying definitions that specify which types of practices are encompassed within terms such as a "lecture." An operationally precise definition of lecturing, for example, would indicate the length of time an instructor must be speaking to be considered a lecture, whether a period of verbal exposition interspersed with questions or other activities constitutes a lecture, and so on (Schonwetter, 1993). This precision is important because within a lecture an instructor may actually be utilizing principles of effective instruction or setting the stage for interactive modalities (Perry & Smart, 1997; Saroyan & Snell, 1997; Schwartz & Bransford, 1999). Indeed, in practice some active learning techniques such as Scientific Teaching do not entirely preclude the use of lecturing but instead aim for only 34% of the class period to be devoted to lecture, ideally broken into segments no longer than 10 minutes (Miller, Pfund, Pribbenow, & Handelsman, 2008). Guided inquiry is an approach where instructors conduct mini-lectures as a way of providing a scaffold to more constructivist activities (Hmelo Silver, Duncan, & Chinn, 2007). Finally, Crouch and Mazur (2001, p. 975) note that in Peer Instruction classrooms, lecturing also takes place: "We typically devote one-third to one-half of class time to ConcepTests and spend the remainder lecturing." In these cases the lecture serves as a way to introduce new topics and to prepare students for more in-depth activities, which highlights the fact that "lecturing" can serve a pedagogically useful purpose.

Ultimately, the lack of carefully specified definitions for teaching methods is an issue because it results in the absence of a shared view of what these methods mean in practice among the research community (Menekse, Stump, Krause, & Chi, 2013). As a result, researchers often fail to articulate what specific behaviors constitute lecturing and other types of teaching. For example, in an experiment comparing lecturing to interactive teaching, the lecturing condition is defined as "using PowerPoint slides to present content and example problems and also (showing) demonstrations" (Deslauriers, Schelew, & Wieman, 2011, p. 862), a definition which encompasses a diverse array range of instructional behaviors and technologies. Perhaps more problematic is the definition provided for the experimental condition (i.e., interactive teaching): "There was no formal lecturing; however, guidance and explanations were provided by the instructor throughout the class" (Deslauriers et al., 2011, p. 863). The exact difference between instances where the instructor provides "guidance and explanations" and is engaged in "formal lecturing" is not provided. Such an approach, in failing to articulate the precise nature of the experimental and control conditions, ultimately raises questions about the validity of the results (Derting et al., 2011; Hora, 2014a), which may be one reason for discrepancies across studies comparing different modes of teaching (e.g., Kirschner, Sweller, & Clark, 2006).

Researchers also commonly describe teaching in binary terms, with mutually exclusive categories such as "active" or "passive" teaching (Menekse et al., 2013) or "lecturing" and "active learning" (Freeman et al., 2014). Such an approach echoes findings from research on faculty cognition that posits instructors' thinking about teaching can be described as either student centered or instructor centered (Kember, 1997). However, this body of research has been critiqued for its lack of ecological validity, as "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). Instead, in practice faculty actually exhibit multiple, often contradictory beliefs about learning that are closely linked to the task at hand (Hora, 2014b). In terms of teaching, this more nuanced view suggests that while some instructors may solely use a "pure" lecturing or active learning approach, there are more subtle variations of these categories in real-world classrooms in terms of time allocated to different types of teaching as well as underlying pedagogical intentions.

To understand the nature of classroom teaching and learning at a finer grain size, Chi (2009) developed the DOLA framework to deconstruct the descriptor "active learning." Indeed, the different active learning traditions noted above (e.g., Scientific Teaching, Peer Instruction) involve distinct approaches to course design and classroom teaching. The DOLA framework was based on the premise that descriptors such as active learning limit the field in a number of ways, including the failure to specify conditions for experimental research in precise terms, and by providing educators with overly ambiguous language to use when attempting to change their teaching from passive to active modalities (Chi & Wylie, 2014). A core feature of the framework is the ICAP hypothesis, which posits that three distinct types of student engagement in the classroom exist in ascending order of efficacy: being active (where students are visibly engaged in activities that activate their own knowledge); being constructive (where students are visibly engaged in activities where they generate their own knowledge); and being interactive (where two or more students are visibly engaged in activities that develop knowledge). The evidentiary base indicates that each of these modalities is more effective than when students are "being passive" in the classroom, but that important differences exist among them as well (Chi, 2009; Chi & Wylie, 2014). This framework is predicated on the notion that while challenging for researchers, these different modalities can be empirically observed in the classroom by documenting the overt behaviors of students. As a result, the DOLA framework represents a significant advance toward the goal of providing researchers with the tools necessary to detect subtle features of active learning in the classroom.

Assuming Student Cognition From Ill-Defined Descriptors of Teaching

From this reliance on reductionist and binary descriptors of teaching arises the second issue: the conflation of specific instructional activities with distinct modes of student cognition (Chi & Wylie, 2014). That is, it is not uncommon to see researchers claim that a particular instructional practice (e.g., small group work) is synonymous with a specific mode of student cognitive engagement such as actively constructing new knowledge. As Chi and Wylie (2014, p. 235) observe, "simply asking students to work together does not automatically make an activity interactive." One of the reasons such an assumption is mistaken is the inherent variability in how different instructors may implement a particular instructional technique. For example, in a study on how the Peer Instruction method (Crouch & Mazur, 2001) was being implemented in undergraduate physics classrooms, Turpen and Finkelstein (2009) found substantial variation in the ways in which questions were asked and in the subsequent learning opportunities afforded to students (see also Zhang & Linn, 2013).

Beyond considerations of instructors' various ways of teaching, however, is the long-standing empirical problem of discerning what is going on in students' minds during a lesson (Nystrand & Gamoran, 1991). Thus, the field of science education are faced with the problem of how to discern the relationship between types of student cognitive engagement and instructional activities without falling prey to the trap of assuming that one always and unequivocally predicts the other. While exploring these dynamics in controlled, lab-based settings is important, so too are studies that examine the relationship between teaching and cognition "in the wild" of actual classrooms. Such research can serve the pressing need to describe and document instructional practice in the nation's postsecondary classrooms (PCAST, 2012) while accounting for the role of local contexts in shaping how students experience different types of teaching. To achieve such fine-grained descriptions, one method is particularly well suited to this task: classroom observations.

Using Classroom Observation Protocols to Study Teaching

Classroom observation is a technique for collecting educational data where researchers take notes and/or code teacher and student behaviors in actual classrooms or from videotaped lessons. As such, classroom observations are part of a larger class of research tools for studying behavior that is used in fields such as cultural anthropology, consumer behavior, and ethology. Classroom observations are used across the educational spectrum from elementary to postsecondary institutions, but the technique is more pervasive and instruments more rigorously tested in K-12 settings, where protocols such as the CLASS protocol (Pianta & Hamre, 2009) and Charlotte Danielson's Framework for Teaching (Danielson, 2013) are central features of teacher evaluation systems in districts across the country. Observations in postsecondary settings are generally used in less high-stakes situations such as professional development and peer mentoring (Chism, 2007). For these applications, the protocols used are often unstructured rubrics where observers take notes but have no prespecified directions about what behaviors or facets of teaching to record and in what fashion.

As interest in the quality and efficacy of postsecondary instruction has increased in recent years, more structured observation protocols have been introduced to the field. The Teaching Behaviors Inventory (TBI) was one of the first widely used protocols in postsecondary settings (e.g., clarity and organization). The TBI is a 60-item instrument composed of eight categories of teaching that requires observers to assign a score on a five-point scale ranging from "almost never observed" to "almost always observed" at the conclusion of the class (Murray, 1983). Another instrument is the Reformed Teaching Observation Protocol (RTOP), which has its basis in the standards-based reform movements in science and math education (MacIssac & Falconer, 2002). The RTOP consists of 25 items scored at the end of a class, which can then be used to classify instructors into one of five categories. Two of these categories represent teacher-centered classrooms (e.g., category one represents "straight lecturing") and three represent learner-centered classrooms (e.g., category five represents "active student involvement in open-ended inquiry") (Ebert-May et al., 2011). Researchers have also developed several other protocols that are similar in structure and intent (e.g., Walkington et al., 2011).

Three characteristics of protocols such as the TBI and RTOP are worth noting. First, in assigning single scores at the end of a class period, these instruments ignore the role of time and the duration with which specific teaching behaviors are observed. Second, while these instruments are designed to capture distinct dimensions of teaching, they capture relatively coarse measures of instruction. Further, the summative scoring procedure makes it impossible to explore the interactions among these dimensions over time. Finally, it is important to note that ultimate purpose of protocols such as the RTOP is to evaluate the quality of instruction rather than to describe it, which is an approach not without limitations.

A recent review of the reliability of evaluative protocols found that ratings of individual teachers varied considerably across analysts (Guarino & Tracy, 2012), and that rater bias (e.g., preexisting beliefs about what constitutes high-quality teaching) is a major reason for the high degree of variation (Cash, Hamre, Pianta, & Meyers, 2012). Partly in response to these concerns, as well as to the perceived need for more rigorous descriptions of practice, researchers have recently developed a class of observation protocols that focuses on describing teaching in fine-grained terms, including the TDOP and the Real-Time Instructor Observing Tool (West, Paul, Webb, & Potter, 2013). These newer, more descriptive approaches are also motivated by a desire to shift the analytic focus from the instructor alone to a more comprehensive and systemic account of the classroom.

The Teaching Dimensions Observation Protocol

The development of the TDOP was largely inspired by a growing movement in educational research that focuses on describing practice as it unfolds in real-world settings, rather than attempting to disseminate best practices with little attention to local contexts (Coburn, 2003; Spillane et al., 2001). Descriptive research of educational practice has been largely motivated by the realization that the implementation of policy as well as the adoption (or rejection) of curricula or teaching methods is strongly shaped by the cultural norms, routines, and structural constraints within particular school settings (Spillane, Reiser, & Reimer, 2002). Practice-oriented scholars have drawn on theories of situated and distributed cognition that emphasize the interdependence of individuals' cognitive processes and the environment in shaping how individuals make decisions, the nature of knowledge, and learning itself (e.g., Brown, Collins, & Duguid, 1989) to study educational practices such as principals' administration of teacher evaluation policies (Halverson & Clifford, 2006), individuals use of mathematics in real-world settings (Lave, 1988), and teachers' decisions about in-class teaching (Schoenfeld, 1999).

In much of this work scholars pay close attention to the role that intentionally designed tools or artifacts (e.g., technology) play in mediating activity (Wertsch, 1991). This mediation process occurs through users' perceptions of limited avenues of potential behaviors (Greeno, 1998) and by artifacts acting as "scaffolds" for learners to perform tasks beyond their existing capacities (Pea, 1993). For example, an individual may perceive that a chair is for sitting, or a teacher may perceive that PowerPoint slides afford the distillation of knowledge into bullet points (Adams, 2006). Ultimately, this perspective emphasizes the interdependence of individuals and their environment in shaping *why* people act the way they do as well as *what* constitutes practice itself.

In a study on how leaders develop professional communities within schools, Halverson (2003, p. 3) built on these ideas by advancing the systems-of-practice approach, which emphasized how networks of tasks, structures, and artifacts (e.g., tools, policies, and procedures) created "complex webs of practice in organizations." Besides the critical role that artifacts play in shaping practice, the systems-of-practice approach also emphasizes the role of time at both the microlevel in terms of task performance and the macrolevel in terms of cultural norms that develop over time as activities are repeated within particular groups and settings. Given its focus on describing educational practice in such a comprehensive manner, my colleagues and I adapted Halverson's (2003) framework and applied it to the study of classroom teaching.

This process began with an instrument originally designed to study inquiry-based science teaching in middle schools that reflected core ideas of the systems-of-practice approach, primarily due to its inclusion of multiple categories to characterize instruction and the innovative use of a time-sampling framework (Osthoff, Clune, Ferrare, Kretchmar, & White, 2008). Using this protocol as a foundation for the TDOP, we identified five key aspects of classroom dynamics: *teaching methods* (e.g., small-group discussion), *pedagogical strategies* used in the classroom (e.g., organization), types of *student–instructor interactions* in the classroom (e.g., types of questions posed), the types of *cognitive engagement* that instructors place on students, and the use of *instructional technology* (e.g., chalkboards). Each category contains several codes that represent specific, overt, and observable behaviors of the instructor and their students in the classroom. 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 (see Table 1).

Two of the categories featured in the TDOP require further examination. First, the teaching methods category includes several codes that encompass instructors' engagement in verbal exposition, which is commonly known as lecturing. In the TDOP, this mode of instruction is decomposed into distinct types of lecturing that implicate other people (e.g., Socractic lecturing) or specific tools and artifacts (e.g., premade visuals such as PowerPoint slides). The latter aspect of lecturing is particularly important given the focus on artifact use as a key mediator of activity in the instructional systems-of-practice framework. Second, the cognitive engagement category refers to the types of cognitive activity that students may be experiencing in the classroom. This category is based on research demonstrating that the type and degree of student cognitive engagement in the classroom is a key aspect of learning (Blumenfeld, Kempler, & Krajcik, 2006; Chi & Wylie, 2014). Measuring cognitive engagement is inherently difficult, and measurement strategies include inferring student engagement from overt student–instructor interactions or in-class learning activities (Nystrand & Gamoran, 1991). Despite the challenges associated with inferring student cognition, it is an important dimension of instruction to capture and is one of the distinguishing features of the TDOP in comparison to other descriptive observation protocols that use a similar time-sampling framework (Smith et al., 2013; West et al., 2013).

While the study reported in this paper is not a validity study for the TDOP, it is important to address issues related to validity and reliability for new research instruments. 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 (Kane, 2001; Porter, 2011). 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, we tested face and construct validity for each of the codes and categories 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 interrater reliability (IRR), which ensures that different analysts will use an instrument in a similar manner across cases. The training procedure for the TDOP is thus rather extensive and places considerable focus on developing IRR. As further development with the TDOP continues, additional validity and reliability evidence will need to be gathered (e.g., test–retest reliability).

METHODS

This study took place at three large, public research universities in the United States and Canada in the spring of 2013. The three study sites had similar undergraduate enrollments and external research funding. We selected these research universities in part because of

TABLE 1
Description of TDOP Categories and Codes

TDOP Category	Code	Description of Code
Teaching methods		
Lecturing	L	Instructor speaks to students with no media
Lecturing with premade visuals	LPV	Instructor speaks with premade visual media (e.g., PowerPoint slides)
Lecturing with hand-made visuals	LHV	Instructor speaks to students with hand-made visuals (e.g., writing on chalkboard)
Lecturing with demonstration	LDEM	Instructor speaks while using demonstrations
Socratic lecture	SOC-L	Instructor speaks while asking questions (two 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
Multimedia	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
Instructor–student interactions		
Rhetorical questions	IRQ	Instructor poses questions without 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 without a known solution

(*Continued*)

TABLE 1
Continued

TDOP Category	Code	Description of Code
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
Clicker response systems	CL	Clicker response systems
Demonstrations	D	Laboratory demonstration equipment
Digital tablet/document camera	DT	Machine used to project images and writing on screen
Movies	M	Movies (e.g., YouTube movies)
Simulations	SI	Graphic simulations and animations

the large number of undergraduates being trained in STEM disciplines at these institutions, as well as the fact that all three had STEM pedagogical improvement initiatives underway at the time of data collection. Personnel active in these initiatives provided initial contacts for our team of researchers.

Faculty were included in the study population if they were listed as instructors in the course schedule for the spring semester. One hundred sixty-five individuals were contacted via email with a request to participate in the study, and 56 ultimately participated (34% response rate). The participants represented the following disciplinary groups: biology ($n = 18$), mechanical engineering ($n = 12$), geology ($n = 15$), and physics ($n = 11$). We selected these disciplines due to the large populations of instructors across the study sites and for their leadership in STEM education initiatives.[5] 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 2). It is important to note that the percentage of instructors not on the tenure-track represented in this study (38%) was roughly in line with the proportion of contingent faculty at participating institutions where data were available (i.e., 33% and 47%).

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.

Training Procedures for the TDOP

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 took approximately 28 hours of training spread out over 2 weeks

[5]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 2
Description of Sample

	Participant (*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

to reach acceptable IRR.[6] The first phase of training in the TDOP involved introducing the raters to each of the codes and rules for applying them in different situations, as well as the process for applying the codes in the classroom. This introductory meeting was followed by two 3-hour long sessions where the entire group coded 10–15 minute segments of videotaped classes, followed by extensive discussions about areas of disagreement while the group also developed a shared understanding of coding rules and procedures. All videotapes used for the training were downloaded from the YouTube channels of research universities. Videotapes for the training were also selected from large physics and biology courses (i.e., "lecture" classes in stadium-style classrooms) to account for variability across disciplines.

Next, the group coded two 50-minute long classes that were followed by formal IRR testing using Cohen's kappa. Cohen's kappa is an 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. The calculations for kappa were made between all possible pairs of raters (i.e., six pairs), with kappa values disaggregated at the code

[6]It should be noted that the claim made in the Smith et al. (2013) paper that adequate training to establish IRR using an abridged version of the TDOP can be conducted in 1.5 hours, runs counter to the experiences of the research group whose work is reported in this paper as well as others experienced in conducting reliability trainings for observation-based research (Joe et al, 2013). While the extensive nature of the training for the TDOP reported in this paper may appear daunting, it is important to recognize that collecting behavioral data in a scientifically rigorous manner is no easy task. As such, drastic reductions in training time made to simplify data collection procedures for researchers should be carefully weighed against potential costs in terms of the quality and utility of the data. That said, our group is actively developing an online training module that will reduce the training time described in this paper.

TABLE 3
Description of TDOP Interrater Reliability Scores for Analysts

Analyst	Teaching Methods	Pedagogical Moves	Interactions	Cognitive Engagement	Instructional Technology
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

category level (e.g., teaching methods) to make it possible to assess raters' agreement on each dimension. It is important to note that the data structure for calculating kappa scores is laid out as a table such that rows represent each code at a particular 2-minute interval and each column is one of two raters. Thus, each row is an "event" indicating whether each rater coded that code as being present in the interval or not. These calculations were all conducted using the TDOP Web site, which has a built-in capability for testing IRR. This process of coding two 50-minute long classes, calculating kappa scores across all rater pairs, examining areas of disagreement, and then returning to another round of coding two new classes took place three times until adequate kappa values were obtained (see Table 3).

Through this intensive training process and formal IRR testing, the reliability of the evidence gathered for the study was demonstrated. In future applications of the TDOP, the training procedures described above are being enhanced by following training protocols used in large-scale observation research. These procedures include testing reliability using a larger number of testing videos (i.e., 5% of total dataset) and testing each rater against precoded videos that are considered to be the "standard," representing perfect application of the instrument as determined by a group of experts (e.g., Joe, Tocci, Holtzman, & Williams, 2013).

Data Collection

After obtaining permission from the instructor, analysts sat near the back of the classroom 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 2-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 may be coded within the same interval. Furthermore, 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 demands 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 training and the fieldwork the team used a form of the TDOP that included 47 codes, but in this paper data are reported for only 33 of these codes. The 14 removed codes were primarily instructional technology codes that were not regularly observed, and their removal allows for a more concise presentation of study results.

Data Analysis

Data from the TDOP instrument were exported from the Web site into spreadsheets where individual 2-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.” These raw data were analyzed by calculating simple proportions of individual codes, predetermined code combinations (i.e., TDOP codes mapped onto the DOLA framework), and social network analyses of data from different course levels (i.e., lower and upper division).

Identifying Frequencies for Individual Codes

First, I report the proportion of times that a particular code was observed across all 2-minute intervals. All figures are rounded to two decimal places. These data are reported for the entire sample as well as by groups that reflect potentially important points of variability within academic contexts including discipline, course level, and class size. It is important to note that analysts scored codes as present if the corresponding practice was observed for any portion of a given 2-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. In addition to reporting simple proportions, I calculated the extent of “straight lecturing” in the study sample, which is defined as periods of verbal exposition (as captured by any of the lecturing codes) where none of the codes signifying active learning are present (e.g., student response [SR], problem solving (PS), students’ novel questions [SNQ], student response (SR), creating (CR), and student peer interactions (PI)). The raw TDOP data were analyzed to categorize instructors based on the maximum length of time they spent teaching in this manner (i.e., up to 20 minutes, 21–40 minutes, over 40 minutes).

Identifying the Prevalence of Active Learning

Next, to identify the prevalence of overt student behaviors, indicative of different types of active learning, codes from the TDOP were mapped onto the three types of active learning discussed in the DOLA framework (Chi & Wylie, 2014). Originally, an attempt was made to distill aspects of well-known active learning strategies such as peer instruction (Crouch & Mazur, 2001) and scientific teaching (Handelsman et al., 2004) that could be linked to TDOP codes, but this approach was rejected because these strategies were not easily operationalized in terms of overt behaviors such that the underlying intent and breadth of the strategy could be captured. Other strategies, such as collapsing descriptive codes into global indicators of student-based activity and instructor-based activity (e.g., Smith et al., 2013), were also not used due to the aforementioned limitations with reducing classroom complexity to such coarse categories as well as the limitations in assuming particular modes of student cognition based on the primary interlocutor in class.

While the TDOP was not developed with either the DOLA framework or other categories of active learning in mind, several TDOP codes focused on student behaviors were well suited to the three categories identified by Chi (2009). The mapping of the two approaches was also made possible by the focus on visible engagement in specific activities emphasized in the DOLA framework. As Chi and Wylie (2014, p. 220) note, while “far from perfect, overt behaviors are a good proxy to reflect different modes of engagement.” The following cross-protocol correspondences were then identified (see Table 4).

TABLE 4
Description of TDOP Code Combinations Used to Capture Elements of the Differentiated Overt Learning Activities (DOLA) Framework (Chi & Wylie, 2014)

DOLA Category	TDOP Category and Code
Being active	Student–instructor interactions: Student response (SR) Or Cognitive engagement: Problem solving (PS)
Being constructive	Student–instructor interactions: Student novel questions (SNQ) Or Cognitive engagement: Creating (CR)
Being interactive	Student–instructor interactions: Student peer interactions (PI) And Cognitive engagement: Creating (CR)

Being Active. The "being active" modality is defined as students being visibly engaged in activities that activate their own knowledge related to course content. Examples from the DOLA framework include following experimental procedures, repeating ideas out loud, highlighting while reading, and copying solutions from chalkboard. To capture the *being active* mode, two TDOP codes were identified (SR or PS). The codes included in this category capture question-and-answer exchanges between students and teachers, with a focus on whether the student responds to an instructor's question. Then, the cognitive engagement mode of problem solving indicates that students are being observed actively engaged in working on closed-ended problems, often (but not always) via desk work (DW).

Being Constructive. The "constructive" modality is defined as students visibly engaged in activities where they generate their own knowledge beyond materials presented in class. Examples from the DOLA framework include self-explaining, drawing concept maps, asking comprehension questions, solving problems requiring the construction of knowledge, and designing a study. To capture the *being constructive* mode, two TDOP codes were identified (SNQ or CR). SNQ refers to instances where students were observed asking original questions to the instructor, as opposed to simply asking for clarification on a topic. The cognitive engagement code of CR refers to instances where instructors give students an open-ended task where no fixed answer exists.

Being Interactive. The "interactive" modality is defined as two or more students visibly engaged activities that develop understanding beyond materials presented in class. Examples from the DOLA framework include working in groups or pairs or interacting with feedback from the instructor. To capture the interactive mode, two TDOP codes were identified that must be observed together within a 2-minute interval (PI and CR) to be included in the *being interactive* category. PI refers to observable interactions between two or more students. Then, because it is possible for PI to occur without any sort of constructive aspect to the work and/or students engagement, the interactive modality is only incurred when PI is observed with the CR mode of cognitive engagement.

It is important to note that the TDOP codes selected for this exercise do not capture the breadth of active learning as posited by the DOLA framework. Future research and development will involve the identification of other overt behaviors that can reliably capture additional aspects of active learning (e.g., student self-explanation, instructor use of varied representations). Also, these codes are not intended as latent constructs that indicate the presence (or absence) of instructional quality, but instead are best viewed as indicators of the three types of active learning as suggested by the DOLA framework. To identify the frequency with which these indicators of active learning were observed, I created one new variable using the "and" operator in SPSS statistical analysis software to capture frequencies for the PI and CR code combination.

Social Network Analysis

Next, 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' 2-minute intervals as rows (mode 1) and TDOP codes as columns (mode 2). Using UCINET (Borgatti, Everett, & Freeman, 2002), the two-mode 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 SGW and PS could have a value of three, which means these two codes were cocoded in three intervals across all instructors in the matrix. I then used the program Netdraw to graph the co-occurrences between each pair of codes across all instructor intervals. The results from the valued co-occurrence matrix were also used to report the strength of ties between active learning indicator codes and other codes.

RESULTS

What Teaching practices are Employed by a Group of Science and Engineering Faculty?

To answer the 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 interconnected and temporal nature of distinct dimensions of instructional practice as suggested by systems-of-practice theory, I present a pair of graphs created using social network analysis techniques.

Frequency of Individual Codes

The data in Table 5 represent the proportion of times that each TDOP code was observed across all 2-minute intervals for the entire sample as well as for each disciplinary group.

These data highlight the prevalence of certain instructional practices across the five TDOP dimensions. Notable results include an extensive amount of lecturing with premade visuals (LPV, 64% of all 2-minute intervals), lecturing with hand-made visuals (LHV, 27%), the administration of assessments (A, 11%), and small group work (SGW, 11%). These results highlight the prevalence of instructional technologies used in the lecturing mode of instruction, particularly that of PowerPoint slides and chalkboards. Then, given the predominance of the "lecturing" descriptor to connote extensive periods of verbal exposition with no student engagement (i.e., a "straight" lecture), and the related goal for faculty to reduce the length such periods of exposition (e.g., Handelsman, Miller, & Pfund, 2007), the

TABLE 5
Classroom Observation Data Using the TDOP by Discipline

Discipline		All	Biology	Mechanical Engineering	Geo-science	Physics
Instructors		56	18	12	15	11
Total 2-minute intervals		2,514	751	527	767	469
Teaching methods						
Lecturing	L	.06	.02	.06	.10	.08
Lecturing with premade visual	LPV	.64	.86	.56	.63	.43
Lecturing with handmade visual	LHV	.27	.11	.52	.14	.49
Lecturing with 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
Multimedia	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
Instructor–student interactions						
Instructor rhetorical questions	IRQ	.12	.10	.10	.12	.17
Instructor display questions[b]	IDQ	.36	.46	.39	.31	.29
Comprehension questions	ICQ	.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
Clicker response system	CL	.10	.16	.08	.04	.14
Demonstrations	D	.02	.01	.05	.01	.04
Digital tablet/document camera	DT	.10	.11	.21	.03	.08
Movies	M	.03	.03	.03	.02	.02
Simulations	SI	.01	0	.02	0	.01

[a] 2^+ questions posed.
[b] Seeking new information.
[c] To instructor question.

data were analyzed to categorize instructors based on the length of time that they used a lecturing modality without any form of visible student engagement, as indicated by the TDOP codes. The results indicate that 61% of the sample ($n = 34$) lectured with no visible student engagement 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. These data indicate that the majority of faculty in the study sample engaged in what is popularly known as a “straight lecture” for relatively brief segments of time within their class. That said, nine instructors were observed teaching in ways where students were passive for almost the entire class period.

Other important aspects of teaching include pedagogical strategies not tied to any particular teaching method including the use of anecdotes and examples (ANEX, 22%) and organizational markers (ORG, 10%). Similarly, faculty employ different approaches to interacting with students, particularly through the posing of questions, including rhetorical questions where students are not expected to answer (IRQ, 12%) or more open-ended questions, known as display questions that solicit specific information from students (IDQ, 36%). Different types of student cognitive engagement were also documented including making connections to the real world (CN, 25%) and a variety of instructional technologies were observed, especially PowerPoint slides (PP, 57%) and chalkboards (CB, 19%).

Variation by Disciplinary Group. This analysis also highlights variations among the disciplinary groups included in the study sample. While an extended analysis of these data is beyond the purview of this paper and has been conducted elsewhere (see Hora & Ferrare, 2013), it is worth pointing out certain differences across groups. For example, in regard to teaching methods, the biologists in the study sample engaged in more lecturing with premade visuals (LPV, 86%) and assessments (A, 17%) than other disciplines, whereas mechanical engineers and physicists used more lecturing with hand-made visuals (LHV, 52% and 49%, respectively), working through problems (WP, 26% and 19%), and desk work (DW, 12% and 12%) than the biologists and geoscientists. Other differences are apparent across other dimensions, including the relatively high incidences of effective pedagogical strategies such as organization (ORG, 12%) and emphasis (EMP, 9%) by biology instructors, the high rates of student comprehension questions (SCQ, 20%), and the use of digital tablets in mechanical engineering classes (DT, 21%). The use of tablets was particularly pronounced at one institution, which had previously invested in upgrading classroom technologies. Similarities across groups are also worth noting, such as the similar rates of certain teaching methods (SGW), instructor-student interactions (PI) and cognitive engagements (PS).

Variation by Course Structure. Given that teaching practices may be influenced by contextual factors such as course level and class size, Table 6 includes data grouped according to these variables. Besides highlighting another point of variability within the data, this analysis also underscores the important role that organizational contexts plays in shaping how instruction unfolds in the classroom as suggested by systems-of-practice theory.

In terms of differences observed according to course level, analysts in lower division classrooms observed less small group work (SGW, 9% and 15%, respectively), peer interactions (PI, 9% and 14%), and creating cognitive engagements (CR, 1% and 5%). Results such as these suggest that upper division courses tend to afford students more opportunities for active engagement with one another and the course material. Variation in teaching practices were also evident in classrooms of differing sizes, though in several cases no clear patterns were evident. For example, lecturing with premade visuals was observed at higher rates in courses with 26 to 100 students (LPV, 71%) and 200 or more students (82%) and at lower rates in courses with less than 25 students (48%) and between 100 and 199 students (55%). In other cases, however, the data suggest that the cognitive engagement of creating is

TABLE 6
Classroom Observation Data Using the TDOP by Discipline and Course Level

		Course Level		Class Size			
		Upper	Lower	<25	26–100	101–199	200+
Instructors		22	34	8	18	18	12
Total 2-minute intervals		1,030	1,477	344	828	821	521
Teaching methods							
Lecturing	L	.07	.06	.10	.06	.08	.02
Lecturing with premade visual	LPV	.60	.68	.48	.71	.55	.82
Lecturing with handmade visual	LHV	.33	.24	.36	.18	.42	.15
Lecturing with 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
Multimedia	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
Instructor–student interactions							
Instructor rhetorical questions	IRQ	.07	.15	.07	.14	.11	.12
Instructor display questions[b]	ICQ	.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	.04	.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
Clicker response systems	CL	.08	.12	.00	.05	.14	.19
Demonstrations	D	.03	.02	.04	.00	.05	.01
Digital tablet/document camera	DT	.12	.09	.00	.06	.12	.19
Movies	M	.04	.02	.00	.04	.03	.02
Simulations	SI	.00	.01	.06	.00	.02	.00

[a]2+ questions posed.
[b]Seeking new information.
[c]To instructor questions.

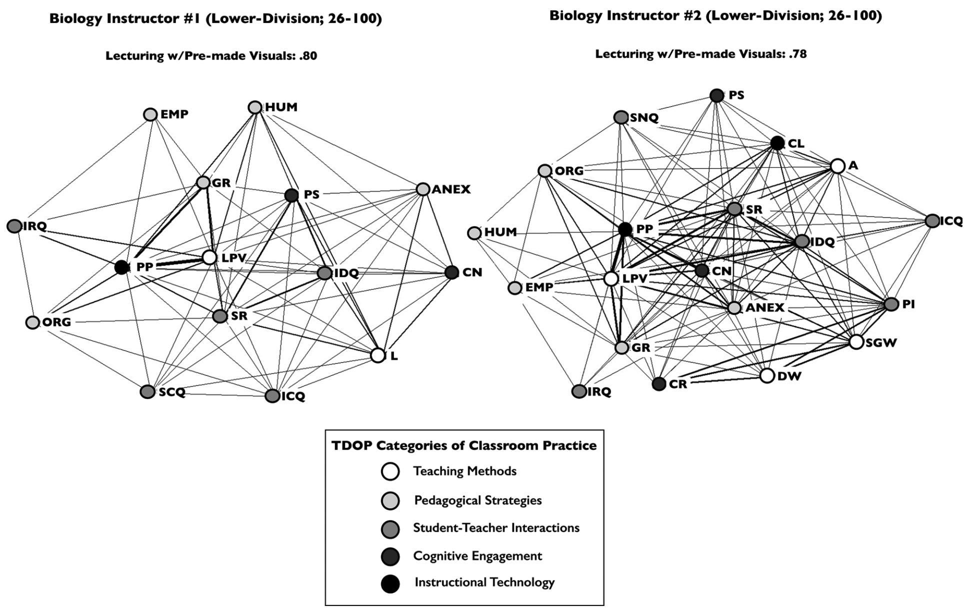

Figure 1. Co-occurence network graph for two biologists teaching similar courses.

most evident in classes with 26–100 students, and that the use of instructional technologies such as PowerPoint slides and clicker response systems becomes more prevalent as class size increases.

Social Network Analysis Graphs

While these data provide a fine-grained account of classroom practice across various settings, the act of decomposing teaching into singular variables such as LPV perpetuates the fiction that teaching can be adequately represented by such measures. Instead, a more accurate approach is to identify the relative intensity with which certain combinations of codes were observed throughout the class period. To illustrate how different aspects of classroom dynamics interact with one another over time, I used social network analysis to depict the TDOP data of two biology instructors at the same institution who were both teaching lower division classes that ranged in size from 26 to 100 students (see Figure 1).

The graphs depicted here illustrate two key points. First, these graphs clearly show that individual codes do not exist in a vacuum but instead often co-occur within the same 2-minute interval. 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 time 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. Ultimately, these images bring to life the advantages of the instructional systems-of-practice framework—how distinct dimensions of instructional practice interact with one another in different ways during the course of a 50-minute class.

Second, the graphs highlight the limitations with conceptions of teaching that collapse the dynamism and complexity of classroom teaching into singular descriptors or variables such as "lecturing." In this case, both faculty exhibited a high degree of lecturing with premade visuals (LPV, 80% and 78%, respectively), yet their overall teaching practices as captured by the TDOP vary considerably. Instructor #1 relied on instructional episodes, which are interconnected modes of instruction that represent a distinct teaching "event" in the classroom, primarily involving lecturing with PowerPoint slides, with other more peripheral behaviors (e.g., using anecdotes) seen less frequently in her class. Instructor #2 also relied on lecturing with PowerPoint slides but demonstrated a far more diversified range of practices that included a variety of teaching methods, pedagogical strategies, and student–instructor interactions. Thus, the graphs demonstrate that while both faculty lectured for extensive periods of time, their teaching approaches and thus the learning environment for their students were quite different, even within similar organizational conditions.

What is the Prevalence and Nature of Active Learning Observed in These Classrooms?

Next, I address the primary question addressed in this paper—to what degree do results indicate the prevalence (or lack thereof) of different types of active learning as measured by categories in the DOLA framework? This question is answered using two different analytic techniques. First, code combinations indicative of active learning are presented according to disciplinary affiliation, course level, and class size. Second, to delve more deeply into how active learning modalities are being used in a specific context, social network graphs are presented that depict instructional practices in upper and lower division courses.

Code Frequencies

The data indicate that the three different categories for active learning suggested by the DOLA framework were observed in varying degrees across the study sample. The "being active" category was most frequently observed, particularly when students responded to instructor questions (SR, 28% of all 2-minute intervals) followed by students engaging in the problem solving modes of cognition (PS, 15%). Much less frequently observed was the "being constructive" category of active learning, which included students asking novel questions (SNQ, 4%) and engaging in creative modes of cognition (CR, 2%), as well as the "being interactive" category of active learning that included both peer instruction and the creating cognitive engagement mode (PI–CR, 2%). In addition, analysts observed variations across disciplinary affiliation and course context (i.e., course level and class size) as seen in Table 7.

Variation by Disciplinary Group. In looking at the results across disciplines, the "being active" category as measured by the TDOP codes of SR and PS was observed the most frequently. The code for student responses to questions was observed the most in biology classrooms (SR, 36% of all 2-minute intervals), followed by physics (29%), geoscience (25%), and then mechanical engineering (21%). The frequencies for problem solving code (PS) was observed in the reverse order. Next, the two categories that have been found to be more effective than the being active modality – that of "being constructive" and "being interactive" – were observed less frequently. The being constructive category involves students either asking novel questions (SNQ) or being actively engaged in open-ended tasks (CR). For both of these codes, the highest rates were in geoscience courses (8% and 4%, respectively). Overall, however, this category of active learning was observed infrequently across the sample relative to the being active category. This was also the

TABLE 7
Description of TDOP Code Combinations Used to Capture Elements of the Differentiated Overt Learning Activities (DOLA) Framework (Chi & Wylie, 2014)

		Disciplinary Group				Course Level		Class Size			
DOLA Category	All Sample	Biology	Physics	Geoscience	Mechanical Engineering	Lower Division	Upper Division	<25	26–100	101–199	200^{+}
Being active											
Student response (SR)	0.28	0.36	0.29	0.25	0.21	0.29	0.28	0.31	0.21	0.30	0.37
Problem-solving (PS)	0.15	0.11	0.17	0.17	0.19	0.14	0.17	0.17	0.14	0.16	0.17
Being constructive											
Student novel question (SNQ)	0.04	0.03	0.02	0.08	0.02	0.05	0.02	0.02	0.04	0.06	0.03
Creating (CR)	0.02	0.03	0.00	0.04	0.01	0.01	0.04	0.00	0.06	0.01	0.00
Being interactive											
Creating and peer interactions (CR and PI)	0.02	0.02	0.00	0.04	0.01	0.01	0.04	0.00	0.05	0.01	0.00

case with the category theorized to be the most effective mode of active learning—that of "being interactive"—which requires both students' active construction of knowledge *and* interaction with peers.

Variation by Course Context. Results for different course levels and class sizes also demonstrate variations in modes of active learning. For the being active mode, the highest rates of students responding to questions were observed in classes with 200 or more students (SR, 37%) and in classes with less than 25 students (31%). This result indicates that either very small or relatively large classes (more than 100) facilitate higher rates of students actively responding to questions, whereas in the mid-size classes (26–100) lower rates were observed (21%). The problem solving mode of cognitive engagement was observed with less variation across course level and class size (from 14 to 17%). Students asking novel questions, one of the indicators for the "being constructive" mode, was observed more frequently in lower division (SNQ, 5%) than in upper division courses (2%), and most often in classes with 101–199 students (6%). These data indicate that students are asking their instructors original questions with greater regularity in lower division courses and in relatively large classes. The other code for this category, that of the creating cognitive engagement, was observed more often in upper division (CR, 4%) than in lower division courses (1%), and most frequently in classes with 26–100 students (6%). Finally, in regard to the "being interactive" category of active learning, the course contexts in which this type of student activity was observed were primarily those of upper division courses (PI-CR, 4%) and in classes with 26–100 students (5%). These results indicate that students are engaging in open-ended problem solving tasks with their peers often in upper division courses and in classes of a particular size (26–100 students).

Social Network Analysis Graphs

Next, I used social network analysis techniques to depict the TDOP data across course levels for illustrative purposes. In the figures presented in this section, a single code representing one of the DOLA categories was the starting point of the analysis (e.g., SR and CR), and all other codes that were observed with these indicators of active learning for at least 10 2-minute intervals were included in the graph (i.e., an ego network). Thus, every code included in these graphs co-occurred with one of the indicators of active learning. A large arrow indicates the primary code of interest in each graph.

Active Learning in Lower Division Courses. To explore the prevalence of active learning in lower division courses, I created two co-occurence network graphs. The first graph highlights the SR code that indicates the "being active" modality (see Figure 2).

This graph is rather complex given that 25 different dimensions of teaching were observed co-occurring with students responding to questions. The codes that most frequently co-occurred with SR in a given 2-minute interval included instructors posing display questions (IDQ, 411 of a total of 1484 intervals), lecturing with premade visuals (LPV, 313 intervals), PowerPoint slides (PP, 287 intervals), and the use of graphics (GR, 166 intervals). The results highlight a commonly observed instructional episode where in the midst of a period of verbal exposition with PowerPoint slides, an instructor would interrupt his or her lecture with one or more questions posed to the class.[7] This suggests that in these courses the

[7]It is important to note that these instructional episodes did not occur identically across all observed class periods, as their exact duration and enactment necessarily varies by instructor and situation.

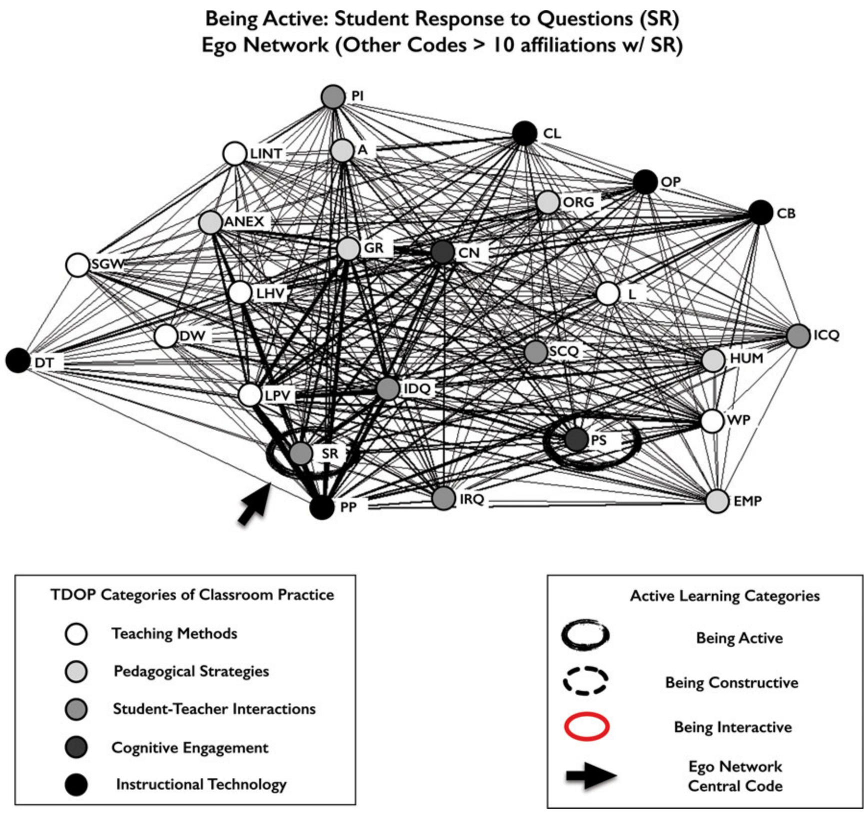

Figure 2. Co-occurence network graph for the "being active" mode in lower division courses as indicated by the "student-response" (SR) code.

"being active" modality as indicated by the SR code is tightly linked to, if not dependent upon, lecturing with specific types of instructional technologies (i.e., PowerPoint slides).

Another graph was created for the other indicator of "being active"—that of PS—but in the interests of space, and since this graph differed from the SR graph only by the exclusion of six codes (L, LPV, SOC-L, ORG, EMP, and IRQ), this graph is not depicted here. The analysis revealed, however, that 19 different dimensions were observed co-occurring with students engaging in PS, with the most frequently observed being instructors' display questions (IDQ, 137 intervals), PowerPoint slides (PP, 108 intervals), peer interactions (PI, 98 intervals), student response to questions (SR, 97 intervals), and small group work (SGW, 95 intervals). These results suggest that the instructional episodes associated with problem solving include questions posed during lectures that feature PowerPoint slides and student responses to these questions either alone or in small groups. Other codes associated with PS that were observed at lower rates of co-occurrence, including clickers (CL, 76 intervals) and desk work (DW, 72 intervals) shed more light on the instructional episodes that involve student problem solving.

The final graph depicts instances where one of the codes signifying the "being constructive" active learning modality—that of students posing novel questions to the instructor (SNQ)—was observed in lower division classrooms (see Figure 3).

The other code for the being constructive mode is not depicted because it was not observed (i.e., creating, CR). Consequently, the "being interactive" modality is not depicted in a graph

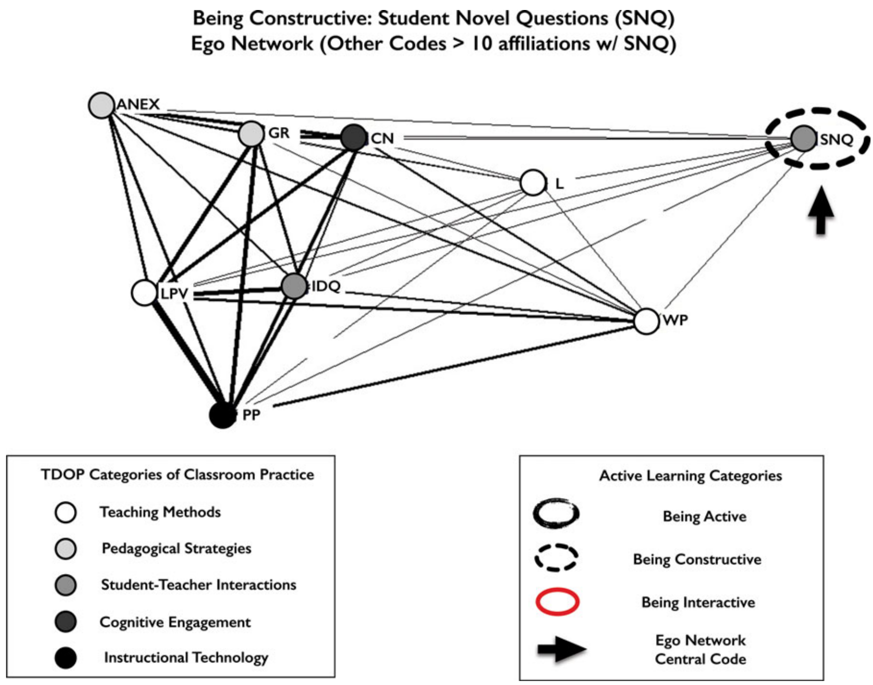

Figure 3. Co-occurence network graph for the "being constructive" mode in lower division courses as indicated by the "student's asking novel questions" (SNQ) code.

for lower division courses. For students posing novel questions, however, eight other TDOP codes were observed with the SNQ code, including lecturing with premade visuals (LPV, 41 intervals), PowerPoint slides (PP, 40 intervals), lecturing (L, 31 intervals), and instructor's working out problems (WP, 14 intervals). It is important to note that in contrast to the preceding graphs, the central code in the ego network is not at the center of the graph but instead occupies a space on the periphery. This indicates that episodes involving SNQ are not frequently observed in lower division classes.

Active Learning in Upper Division Courses. Next, I report findings from the analysis of active learning modalities observed in upper division courses. The first graph highlights the SR code that indicates the "being active" modality (see Figure 4).

This graph depicts the 26 codes that co-occurred with SR. The centrality of the SR code is immediately apparent in the graph, with several thick, dark lines connecting it to other dimensions of instruction. Additionally, a cluster of other active learning modalities is evident at the right of the graph (i.e., being interactive and being constructive), suggesting that a more peripheral set of practices exists at this course level in addition to the core modality of being active. The codes that most frequently co-occurred with SR in a given 2-minute interval included instructors posing display questions (IDQ, 263 of a total of 1030 intervals), lecturing with premade visuals (LPV, 184 intervals), PowerPoint slides (PP, 148 intervals), and the use of graphics (GR, 109 intervals). The results are identical to incidences of SR in lower division courses, albeit with different rates of co-occurrence. This similarity suggests that instructors across course levels commonly interrupt periods of verbal exposition with PowerPoint slides by posing one or more questions to the class.

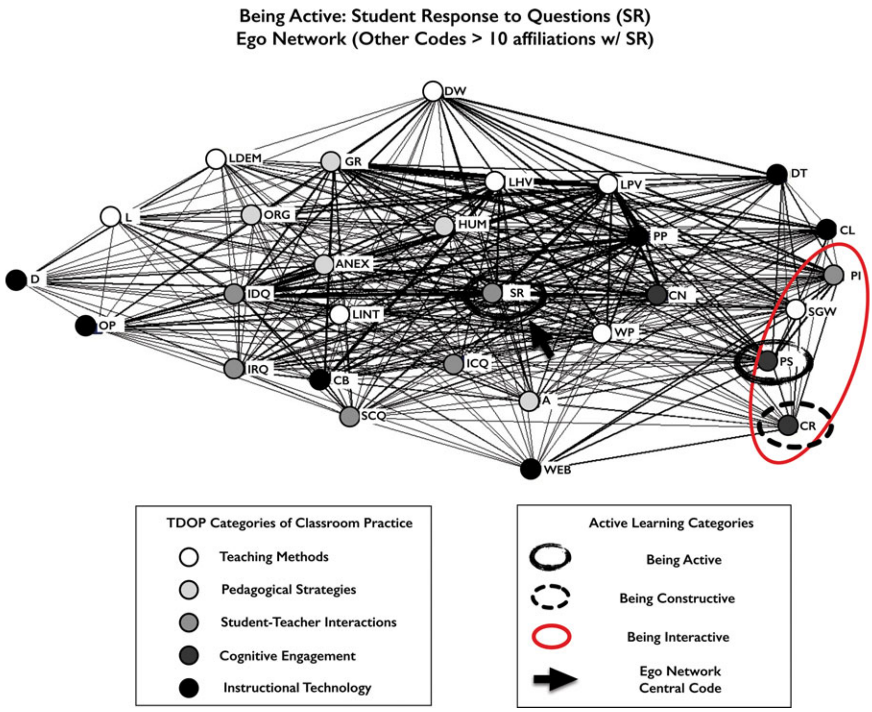

Figure 4. Co-occurence network graph for the “being active” mode in upper division courses as indicated by the “student-response” (SR) code.

In the interests of space, the graph for the upper division PS code, also indicative of the “being active” modality, is not included in this paper. This analysis did indicate that the PS graph only differed from the SR graph by the exclusion of eight codes (LDEM, SOC-L, IRQ, ICQ, SCQ, CN, OP, and D) and the presence of two codes (CR and WEB) not in the previous graph. The analysis revealed 20 different dimensions observed co-occurring with students engaging in PS, with the most frequently observed being small group work (SGW, 127 intervals), student peer interactions (PI, 110 intervals), instructor’s display questions (IDQ, 105 intervals), and lecturing with premade visuals (LPV, 85 intervals). These results suggest that the instructional episodes associated with PS at the upper division level include students working in small groups with their peers on questions posed by the instructor. The instructional episode that links PS to SGW and IDQ as described for lower division courses is also evident here because SGW and IDQ also regularly co-occur (86 intervals). Furthermore, it is notable that many of the questions that spark PS are delivered from a lecturing context—that is, a period of lecturing often preceded instances of student problem solving.

The next figure includes two smaller graphs that depict both the “being constructive” and the “being interactive” modalities (see Figure 5).

In the first graph on the left, one of the indicators for the “being constructive” modality—that of student’s posing novel questions (SNQ)—is depicted as the primary code in the graph. In this graph, only three other codes were observed co-occurring with students posing novel questions, with the most frequently observed being lecturing with premade visuals (LPV, 19 intervals), PowerPoint slides (PP, 13 intervals), and the use of graphics (GR, 10 intervals).

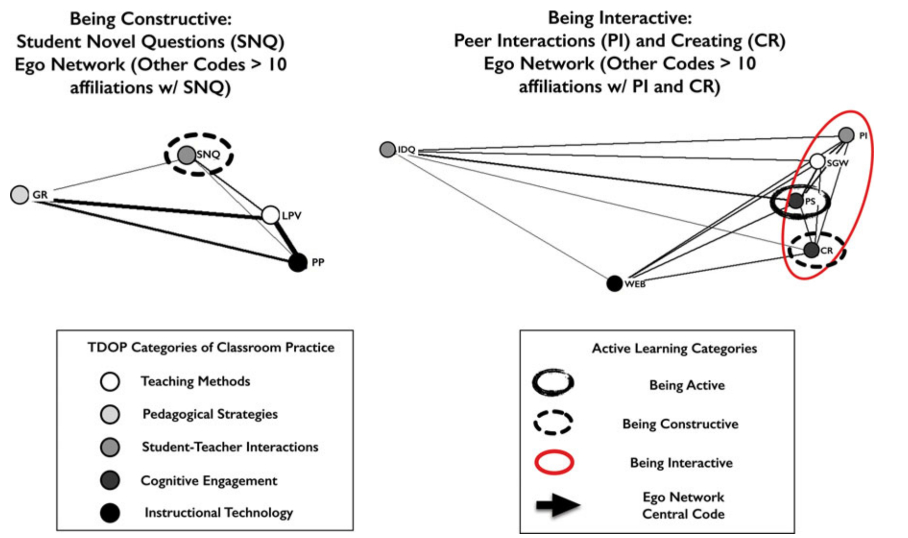

Figure 5. Co-occurence network graph for the "being constructive" mode in upper division courses as indicated by the "student's asking novel questions" (SNQ) code and the "being interactive" mode as indicated by the "peer interaction" (PI) and "creating" (CR) codes.

These results are similar to those observed in lower division courses, where students were observed interrupting instructor's Powerpoint lectures with novel questions.

In the second graph on the right side of Figure 5, the two indicators comprising the "being interactive" modality are shown. No graph is shown for the creating (CR) mode on its own, which is an indicator for "being constructive," because the networks for both CR alone and CR and PI together are identical. This co-occurence network graph depicts all TDOP codes that co-occurred with both CR and PI in the same 2-minute interval. These codes included small group work (SGW, 127 intervals with PI and 40 intervals with CR), instructors posing display questions (IDQ, 67 intervals with PI and 12 with CR), problem solving (PS, 110 intervals with PI and 31 with CR), and Web site use in the classroom (WEB, 33 intervals with PI and 32 with CR). These results show instructional episodes that were observed in classes where instructors posed display questions often focused on open-ended tasks or problems that initiated small group-work involving peer-to-peer interactions. In one case, students were directed to use a Web site as part of a problem solving activity.

DISCUSSION

The overall goal of this study was to advance an approach for studying teaching in naturalistic settings, with the specific aim of demonstrating how active learning modalities can be documented and described in postsecondary classrooms. In this paper, the "problem" of identifying whether active learning techniques are being used, which is of great interest to policymakers and researchers alike (e.g., PCAST, 2012; AAAS, 2012), was reframed from one that involved identifying the prevalence (or absence) of coarse descriptions of teaching to a more nuanced and comprehensive view of instructional practice. The latter formulation of the problem represents an entirely different proposition than the former. Instead of assuming that the fundamental question of understanding the relationship between teaching and learning in *all* settings (both controlled and naturalistic) was a settled empirical question, such that the mere identification of "lecturing" would capture the entirety of an instructor's practice and subsequent student engagement (Freeman et al., 2014; Wieman, 2014), the problem is rather one of first attempting to describe the nature of teaching itself as a scientific problem on its own merit. Indeed, by demonstrating that modes of active learning are often embedded within PowerPoint lectures and that small group work exercises are not synonymous with constructivist activities, the evidence presented in this paper as well as by other researchers (e.g., Chi & Wylie, 2014) suggests that the research program of understanding teaching as a complex behavioral phenomena in its own right is in the early stages. As such, the field of science education would benefit from research instruments that can detect, with precision, what actually happens in real-world classrooms, something that existing surveys and classroom observation protocols are not designed to do. In the remainder of this paper, I elaborate on key findings from the observation of 56 science and engineering instructors using the TDOP, and outline implications for research, practice, and policy that follow from the results.

Prevalence of Instructional Practices in General: The Central Role of Lecturing

First, the results shed light on the types of instructional practices that 56 science and engineering instructors are using in their courses. In terms of single TDOP codes, the most frequently observed included those pertaining to different forms of verbal exposition (i.e., lecturing), particularly in conjunction with premade visuals (LPV, observed in 64% of all 2-minute intervals) and handmade visuals (LHV, 27%). While caution must be taken when isolating these data from their role as part of an interactive system of tools and practices, they do shed light on the prevalence of individual aspects of classroom dynamics. For example, the data speak to the widespread use of particular types of representations and media (e.g., premade visuals such as PowerPoint slides) that faculty use as part of their lecturing.

Specifying distinct types of lecturing according to the different representational modes used in the classroom is also important because such artifacts mediate learners' experiences with course material and the disciplinary processes and communities in which they are embedded (Brown et al., 1989; Lattuca, 2002). The tools and technologies used while lecturing also effectively frame and organize knowledge in particular ways that offer viewers a circumscribed range of potential responses or actions (Greeno, 1998), and which may ultimately reorganize observers' mental functioning (Hawkins & Pea, 1987). While the data reported in this paper are not designed to advance claims on the types of cognitive activities afforded by viewing a series of PowerPoint slides relative to writing on a chalkboard, research on relationships among the use of varied representations, technology, and student learning should be incorporated into debates about instructional reform (e.g., Mayer, 2011).

Furthermore, because 34 instructors (61%) lectured for relatively brief periods (less than 20 minutes), the data indicate that a majority of the faculty in the study were not engaged in extensive lecturing that lacked any interactions with students. Indeed, some instructors also interspersed questions, small group work, and other activities throughout their "lectures." This particular approach to lecturing was evident in the case of the biology instructor featured in Figure 1. Ultimately, these data demonstrate that in practice, lecturing can be a rather complex teaching practice that does not conform to the oft-cited caricature whereby the instructor speaks, with no effort or opportunity for student engagement or activity, for an entire class period. While nine instructors in the study certainly did enact the common notion of the "straight lecturer" as a teacher-centered performance with little care or attention toward engaging students, most did not. Indeed, in some cases, it appeared that lecturing was used as a pedagogical device to support other, more engaged modes of instruction.

It also cannot be immediately assumed that the dominance of lecturing automatically leads to inferior instruction and poor student learning (see Schwartz & Bransford, 1998; Crouch & Mazur, 2001; Miller et al., 2008). Thus, the pertinent question becomes one of the amount and timing of lecturing and whether an underlying pedagogical rationale guides its use. One of the key questions facing the field, then, is how to use these "times for telling" in a way that is pedagogically sound, well executed, and serves to set up robust active learning modalities (Schwartz & Bransford, 1998). That said, while the specific duration and type of lecturing that is the most pedagogically beneficial remains an open question, the weight of the evidence regarding the importance of active learning does suggest that there is room for greater incorporation of nonlecture activities in some classrooms—particularly in the case of those nine faculty who lectured for over 41 minutes with nary a question or activity.

Finally, 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 premade visuals using the TDOP instrument (observed in 64% of all 2-minute intervals), but varies from results reported here regarding the prevalence of long periods of lecturing (e.g., 40^{+} minutes). This result 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.

Prevalence of Active Learning Modalities: Variation Within Instructional Activities

Second, the results also shed light on the prevalence of active learning in the postsecondary classroom. Using indicators of active, constructive, and interactive types of instructional activities as posited by the DOLA framework (Chi & Wylie, 2014), the evidence shows that the most commonly observed type of active learning involved the "being active" category. This most often involved students responding to instructor's questions (SR, observed in 28% of all 2-minute intervals) followed by students engaging in problem solving modes of cognition (PS, 15%). These types of active learning frequently involved instructional episodes that included instructors posing questions that sought new information, either verbally or through electronic means (e.g., clicker response systems), often during a period of lecturing with PowerPoint slides. It is important to note that this category of active learning is at the lowest end of the DOLA framework in terms of efficacy in supporting student learning, and while it is more effective than a lesson where students are completely passive, it is less effective than instruction that facilitates students to be "constructive" or "interactive" (Chi & Wylie, 2014). Ultimately, the results suggest that the "being active" mode is a common instructional approach in postsecondary classrooms, achieved primarily through questions posed to students in the midst of a PowerPoint lecture, whereas other types of active learning are far less common.

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 (CL, 8%) and small group work (SGW, 11%) capture key facets of the strategy and similar to findings in other studies. What is most evident from the comparison of these survey results to the findings reported in this paper is the fundamentally different perspective on teaching offered by the instructional systems-of-practice framework and most survey instruments. Instead of distilling instruction in general and active learning in particular down to single descriptors, which are often assumed to be synonymous with high-quality student cognitive engagement, active learning modalities are seen as being composed of a configuration of distinct yet interrelated dimensions of teaching. The limited capacity of surveys to capture the prevalence (and quality) of active learning is perhaps most evident in the example of small group work, a widely encouraged classroom activity (e.g., Handelsman et al., 2004; PCAST, 2012). However, it cannot be assumed that asking students to work in groups automatically translates into an interactive and effective learning experience (Chi & Wylie, 2014).

The data reported in this paper support this idea by revealing that small group work activities vary in terms of associated cognitive demands between lower division and upper division classes. In lower division classes, small group work was more often associated with closed-ended problem solving modes of cognition (PS, observed with SGW in 95 of 1477 total 2-minute intervals) than open-ended creative modes of cognition (CR, observed with SGW in 0 of 1477). Similarly, for upper division classes, small group work was more often associated with problem solving (PS, observed with SGW in 127 of 1030) than creative modes of cognition (CR, observed with SGW in 40 of 1030). These findings are consistent with prior research that small group activities can be implemented in very different ways (Turpen & Finkelstein, 2009), and highlight the fact that without differentiating among types of student cognitive engagement (e.g., CR and PS), relying solely on whether or not small group work is taking place results in an incomplete and potentially misleading account of active learning in the classroom (e.g., Smith et al., 2013).

Insights Into Variations by Disciplinary Group and Course Characteristics

Finally, the results indicate variations in both teaching practices in general and active learning modalities in particular across different disciplinary groups and course contexts (i.e., course level and class size). While an extended analysis of the nature of and implications for these differences are beyond the purview of this paper, it is worth noting that such variations are consistent with a theory of practice that emphasizes the dynamic interactions between context and activity (Halverson, 2003; Spillane et al., 2001). In particular, the interactions between organizational structure and teaching practices are evident in the different rates of the "being constructive" modality, where students were most frequently observed asking novel questions in lower division (SNQ, observed in 5% of all 2 minute intervals) courses, and in classes with 101–199 students (SNQ, 6%). These and other points of variation suggest that the level of the course as well as structural affordances in the classroom itself may play an important role in shaping both the instructional decisions made by instructors and students' subsequent cognitive activity.

Limitations

There are several limitations to the TDOP and the data reported in this study. 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 student learning) are needed 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 limitation is less problematic with regard 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 was 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 absence of data on laboratory or discussion sections.

Implications for Research, Policy, and Practice

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 teaching in terms of research, policy, and practice.

Research

In attempting to study, understand, and measure classroom teaching as a behavioral phenomenon itself it is clear that singular, decontextualized metrics and binary categorization schemes are insufficient. If the field were to think of the problem in terms of biological classification, it is as if researchers were sometimes focusing solely at the level of classes, while ignoring taxa such as genera or species that would reveal more fine-grained and subtle distinctions among different types of instruction. The instructional systems-of-practice approach and the TDOP represent a step in this direction, but future research should continue to identify, in more precise terms than at present, the types of teaching practices, active learning modalities, and student behaviors that occur in real-world classrooms. Towards this end, my research group is actively field-testing new sets of TDOP codes to capture more nuanced features of active learning techniques as well as instruction that conveys what are known as "21st century competencies" important for life and work (e.g., self-regulation) (Pellegrino & Hilton, 2012). Indeed, some researchers are actively using the basic architecture of the TDOP to develop new methods for studying aspects of active learning in the classroom (Lund et al., 2015).

In addition, researchers should focus on better understanding student behavior and subsequent cognitive demands. 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)

Additionally, in studying whether and how instructors are adopting active learning techniques, researchers should expand their inquiries beyond the classroom. Factors such as careful attention to course design as a whole and how instruction, assessment, and student assignments can work together in a complementary manner to provide students with a rich learning experience are just as important to the provision of high-quality learning opportunities (Freeman, Haak, &Wenderoth, 2011). Furthermore, given increasing evidence regarding which study strategies are most effective in facilitating student learning (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013), researchers should pay as much attention to student study habits "in the wild" as is currently being focused on what faculty do in the classroom.

Practice

One of the most promising uses for descriptive data is that of supporting professional development efforts. A critical feature of professional development is the provision of credible and detailed feedback for instructors that can spark critical reflection (Chism, 2007). Yet data about teaching, beyond ubiquitous end-of-term student evaluations, are often in short supply in many colleges and universities, and faculty are often left with little data upon which to evaluate their teaching effectiveness (Gormally, Evans, & Brickman, 2014). Such reflection is a cornerstone to the ongoing development of professional expertise (Schön, 1983), and I suggest that the TDOP can be useful in providing data for these purposes, especially because faculty may be more responsive to and less threatened by the results in contrast to evaluative protocols such as the RTOP (Yon, Burnap, & Kohut, 2002). This thesis is currently being tested in a field-based study on faculty reactions to data from the TDOP and student evaluations as a new type of formative feedback (see http://tpdm.wceruw.org).

Policy

Unfortunately, the growing evidence supporting the effectiveness of active learning approaches and the binary categorization scheme used to characterize teaching has coalesced into pronouncements that lecturing is the "pedagogical equivalent of bloodletting" and is synonymous with an "inferior education" (Weiman, 2014, p. 8320). As such, the current rhetoric perpetuates the mistaken notion that the term "lecturing" refers to a distinct type of instructional practice that is unequivocally indicative of a specific type of student cognitive engagement, instead of one that may vary in a multitude of ways and actually be used in conjunction with active learning techniques. Such an approach can easily alienate instructors "in the field," as practitioners grow to resent the perceived imposition of particular philosophies or opinions about what constitutes effective or high-quality teaching from external authorities (Henderson & Dancy, 2008; Yon et al., 2002). Indeed, extensive 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, is closely aligned with the existing practices, cultural traditions, and beliefs of the population, instead of being in direct opposition to them (Rogers, 1995; Spillane et al., 2002).

Ultimately, I speculate that reforms that promote active learning 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 lecturing versus interactive teaching framework 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, 1993). Thus, it may not be a matter of eliminating lecture from one's pedagogical toolkit, but instead altering one's lecturing approach to have more of a deliberate pedagogical purpose that engages students in their own learning. 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.

I would like to thank the anonymous reviewers and Jana Bouwma-Gearhart for their feedback on earlier versions of this paper. The research reported here was supported by the National Science Foundation under award DUE#1224624. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Adams, C. (2006). PowerPoint, habits of mind, and classroom culture. Journal of Curriculum Studies, 38(4), 389–411.

American Association for the Advancement of Science. (2012). Describing and measuring undergraduate STEM teaching practices: A report from a national meeting on the measurement of undergraduate STEM teaching, December 17–19, 2012. Washington, DC: Author.

Blumenfeld, P., Kempler, T., & Krajcik, J. (2006). Motivation and cognitive engagement in learning environments (pp. 475–488). New York, NY: Cambridge University Press.

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

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

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational Researcher, 18(1), 32–42.

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.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. Topics in Cognitive Science, 1(1), 73–105.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. Educational Psychologist, 49(4), 219–243.

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

Clark, R. M., Norman, B. A., & Besterfield-Sacre, M. (2014). Preliminary experiences with "flipping" a facility layout/material handling course. In Y. Guan & H. Liao (Eds.), Proceedings of the 2014 Industrial and Systems Engineering Research Conference, May 28–June 3. Montreal, Canada.

Coburn, C. E. (2001). Collective sense making about reading: How teachers mediate reading policy in their professional communities. Educational Evaluation and Policy Analysis, 23(2), 145–170.

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. Educational Researcher, 32(6), 3–12.

Code, W., Piccolo, C., Kohler, D., & MacLean, M. (2014). Teaching methods comparison in a large calculus class. ZDM Mathematics Education 46(4), 1–13.

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.

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.

Duch, B. J., Groh, S. E., & Allen, D. E. (Eds.). (2001). The power of problem-based learning: A practical" how to" for teaching undergraduate courses in any discipline. Sterling, VA: Stylus Publishing.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. Psychological Science in the Public Interest, 14(1), 4–58.

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.

Finelli, C. J., Daly, S. R., & Richardson, K. M. (2014). Bridging the research to practice gap: Designing an institutional change plan using local evidence. Journal of Engineering Education, 103(2), 331–361.

Franklin, S. V., & Chapman, T. (2012). Diversity of faculty practice in workshop classrooms (Vol. 1513, pp. 130–133). Philadelphia, PA: Physics Education Research Conference.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., et al. (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.

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

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.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. American Psychologist, 53(1), 5–26.

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

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

Halverson, R. R., & Clifford, M. A. (2006). Evaluation in the wild: A distributed cognition perspective on teacher assessment. Educational Administration Quarterly, 42(4), 578–619.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., et al. (2004). Scientific teaching. Science, 304(5670), 521–522.

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

Hawkins, J., & Pea, R. D. (1987). Tools for bridging the cultures of everyday and scientific thinking. Journal of Research in Science Teaching, 24(4), 291–307.

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.

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.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). Educational Psychologist, 42(2), 99–107.

Hora, M. T. (2014a). Limitations in experimental design mean that the jury is still out on lecturing. Proceedings of the National Academy of Sciences, 111(30), 3024.

Hora, M. T. (2014b). Exploring faculty beliefs about student learning and their role in instructional decision-making. The Review of Higher Education, 38(1), 37–70.

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

Hora, M. T., & Ferrare, J. (2014). Re-measuring postsecondary teaching: How singular categories of instruction obfuscate the multiple dimensions of classroom practice. Journal of College Science Teaching, 43(3), 36–41.

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.

Joe, J. N., Tocci, C. M., Holtzman, S. L., & Williams, J. C. (2013). Foundations of observation: Considerations for developing a classroom observation system that helps districts achieve consistent and accurate scores. MET project policy and practice brief. Bill & Melinda Gates Foundation.

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

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.

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

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41(2), 75–86.

Lattuca, L. R. (2002). Learning interdisciplinarity: Sociocultural perspectives on academic work. The Journal of Higher Education, 73(6), 711–739.

Lave, J. (1988). Cognition in practice: Mind, mathematics and culture in everyday life. Cambridge, UK: Cambridge University Press.

Lund, T. J., Pilarz, M., Velasco, J. B., Chakraverty, D., Rosploch, K., Undersander, M., & Stains, M. (2015). The best of both worlds: Building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. CBE-Life Sciences Education, 14(2), 1–12.

MacIsaac, D., & Falconer, K. (2002). Reforming physics instruction via RTOP. The Physics Instructor, 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.

Mayer, R. E. (2011). Instruction based on visualizations. In R. E. Mayer & P. A. Alexander (Eds.), Handbook of research on learning and instruction (pp. 427–445). New York: Routledge.

Mazur, E. (2009). Farewell, lecture. Science, 323(5910), 50–51.

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. (2013). Differentiated overt. Learning activities for effective instruction in engineering classrooms. Journal of Engineering Education, 102(3), 346–374.

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.

National Research Council. (2012). Discipline-based education research: Understanding and improving learning in undergraduate science and engineering. Washington, DC: National Academies Press

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., & 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.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), Distributed cognitions (pp. 47–87). New York: Cambridge University Press.

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 instructors' 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. Washington, DC: Executive Office of the President.

Rogers, E. M. (1995). Diffusion of innovations (4th ed.). New York: Simon & Schuster.

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: Basic Books.

Schoenfeld, A. H. (1999). Models of the teaching process. The Journal of Mathematical Behavior, 18(3), 243–261.

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

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. Cognition and Instruction, 16(4), 475–522.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. Educational Researcher, 31(7), 15–21.

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.
Walkington, C., Arora, P., Ihorn, S., Gordon, J., Walker, M., Abraham, L., et al. (2011). Development of the UTeach observation protocol: A classroom observation instrument to evaluate mathematics and science teachers from the UTeach preparation program (UTeach Technical Report 2011–01). Austin: UTeach Natural Sciences, University of Texas at Austin.
Wertsch, J. V. (1991). Voices of the mind: A sociocultural approach to mediated action. Cambridge, MA: Harvard University Press.
West, E. A., Paul, C. A., Webb, D., & Potter, W. H. (2013). Variation of instructor-student interactions in an introductory interactive physics course. Physical Review Special Topics-Physics Education Research, 9(1), 010109.
Wieman, C. E. (2014). Large-scale comparison of science teaching sends clear message. Proceedings of the National Academy of Science, 111(23), 8319–8320.
Yon, M., Burnap, C., & Kohut, G. (2002). Evidence of effective teaching: Perceptions of peer reviewers. College Teaching, 50(3), 104–110.
Zhang, Z. H., & Linn, M. C. (2013). Learning from chemical visualizations: Comparing generation and selection. International Journal of Science Education, 35(13), 2174–2197.