Remeasuring Postsecondary Teaching: How Singular Categories of Instruction Obscure the Multiple Dimensions of Classroom Practice

By Matthew T. Hora and Joseph J. Ferrare

Analyses of postsecondary teaching are limited by overly blunt measures that focus on regularly used teaching methods (e.g., lecturing) while ignoring other important dimensions of classroom practice. This is important because these flawed measures of teaching are being used as a central feature of the national discourse on science and math education. In this article we introduce a new approach to studying postsecondary teaching that captures five distinct dimensions of teaching practice as they interact over time. Using the Teaching Dimensions Observation Protocol, we observed 58 science and math faculty teaching undergraduate courses at three large, public research universities. Results are reported by institution and discipline, with in-depth analyses of two biology instructors that reveals substantial variation between individuals who regularly lecture. Improved measures of postsecondary teaching can be used to interject more empirically sound accounts of teaching into the national debate on science and math education.

How faculty teach in the classroom is commonly measured by questionnaires that elicit the regularity with which someone uses a particular teaching method, such as lecturing or small group work (e.g., DeAngelo, Hurtado, Pryor, Kelly, & Santos, 2009). This approach has contributed to a widespread notion that measuring the self-reported teaching methods that an instructor uses is an adequate and accurate measure of teaching. Not surprisingly, however, evidence from research on K–12 and postsecondary education indicates that educational practice is in fact far more complex than a single descriptor can capture (Cohen & Ball, 1999; Murray, 1983; Spillane, Halverson, & Diamond, 2001). Reducing how an instructor teaches Biology 101, for example, to a variable such as "lecturing" masks other important dimensions of her teaching, such as the types and frequency of questions posed to students or how students are socially and cognitively engaged in the classroom (Hora & Ferrare, 2013). Indeed, research indicates that faculty use of the lecturing method does exhibit important variations such that distinct "lecturing styles" can be identified (Saroyan & Snell, 1997). In taxonomic terms, naming all instances of instructors' verbally delivering material as lecturing equates to conceptualizing and measuring classroom teaching solely at the level of biological classes, while ignoring taxa such as orders or genera that would reveal subtle distinctions among different types of teaching. Although we are not suggesting that teaching can be classified in the same way as organisms, we argue that it is possible to capture a more detailed and rigorous account of instruction than is possible with currently available conceptualizations of teaching and related research instruments.

Given these limitations in how teaching is currently studied, the field of science education faces a critical problem of measurement that has significant implications for policy and practice. In attempting to categorize teaching into simple dichotomous groups (e.g., lecturing vs. interactivity), policy makers are provided with an inaccurate and coarsely grained perspective of teaching that does not reflect the realities of classroom practice. In its 2012 report, the President's Council of Advisors on Science and Technology (PCAST) recommended that the federal government encourage the widespread adoption of "active learning" teaching methods, which are directly contrasted to "a sole reliance on lecturing" (PCAST, 2012). This sentiment is echoed in much of the educational literature, with faculty strongly encouraged to eschew the lecture in favor of other, more interactive techniques (Handelsman et al., 2004; Mazur, 1997), and lecturing being used as a dependent variable with which to distinguish between control and experimental conditions (Deslauriers, Schelew, & Weiman, 2011). Although we do not disagree with the efficacy of many interactive teaching approaches, we suggest that by pitting two higher order measures of teaching (e.g., lecturing vs. interactive methods), observers are glossing over important distinctions and subtleties that make up classroom teaching. For example, instructors will also frequently use a variety of methods and strategies within a single class period, such that a single descriptor necessarily ignores how long particular methods are used and how different techniques are used in combination with one another.

The particular focus on lecturing is also limited by the assumption that any given teaching method is inherently effective or ineffective, regardless of the manner in which it is used in the classroom. This assumption is not supported by empirical research on how faculty implement certain teaching strategies in the classroom. For example, Turpen and Finkelstein (2009) found that a group of physicists implemented the peer instruction technique with a high degree of variability with resulting consequences for the pedagogical quality of the lessons. This finding underscores the importance of understanding how instructors are using particular teaching strategies and behaviors in the classroom with as much precision as possible and how these practices ultimately influence student learning outcomes.

An important source of the reductive conceptualization of postsecondary teaching stems from the way these practices are measured by researchers. Although a variety of classroom observation protocols are available, many are limited by the reliance on open-ended response items that preclude reliability tests or comparison across individuals, a singular focus on the use of teaching methods, or the a priori equation of instructional quality with certain teaching methods. This last point regarding the methodological limitations of protocols that require observers to not only describe but also to evaluate the quality of teaching is particularly important, given the widespread use of such protocols in postsecondary settings. A recent review of reliability of evaluative protocols used in K–12 settings found that ratings varied considerably (Guarino & Stacy, 2012) and also that rater bias (i.e., preexisting beliefs about what constitutes high-quality teaching) is a major reason for the high degree of variability observed in the use of these protocols (Cash, Hamre, Pianta, & Meyers, 2012).

In this article we report findings from classroom observations of 58 math and science faculty from the spring of 2012 using the newly developed Teaching Dimensions Observation Protocol (TDOP) that addresses these methodological limitations while also capturing more subtle and dynamic features of practice than an exclusive focus on teaching methods. An extensive discussion of the conceptual origins of this approach and findings from data collected in the spring of 2010 using a previous version of the TDOP are included in another article (Hora & Ferrare, 2013). The data reported in this article are based on a revised version of the protocol and were collected at three public research universities; all faculty self-selected into the study. The resulting data demonstrate that postsecondary teaching is more complex and nuanced than is suggested through a reliance on descriptors such as "lecturing."

Teaching Dimensions Observation Protocol

The TDOP captures five different dimensions of teaching practice: teaching methods, pedagogical strategies, student–teacher interactions, cognitive engagement, and use of instructional technology. Besides observable teaching methods such as lecturing or small group work, instructors use more subtle pedagogical strategies such as humor, illustrations or anecdotes, and verbally marking transitions between topics that play critical roles in instruction (Perry & Smart, 1997). A singular focus on instructors' behaviors also obscures the critical role that students play in the classroom, which can be measured by observing the nature of questions instructors pose to their students as well as inferences regarding the potential for different forms of cognitive engagement among students (Morell, 2004; Porter, 2002). Finally, capturing how faculty use instructional technology is critical given the increasingly important role that technology plays in teaching and learning.

Within each dimension there exist several detailed codes that observers capture at 2-minute intervals (Table 1). These codes were developed in coordination with teams of math and science faculty and were refined through extensive field testing to enhance face validity of the instrument. Analysts also undertook extensive training to establish interrater reliability (IRR). The following are results of the IRR using Cohen's Kappa for each pair of raters (averaged across the five categories): Analyst 1/Analyst 2 (.783); Analyst 1/Analyst 3 (.759); and Analyst 2/Analyst 3 (.749). It is important to note that several codes in the protocol were revised after data collection at the first institution, and subsequent retraining led to an improved IRR. These revisions pertained to three codes in the cognitive engagement category and one code in the student–teacher interaction category. When data are reported for

TABLE 1

Proportion of 2-minute intervals that each teaching practice was observed.

	Entire sample (*n* =58)	Math (*n* =14)	Biology (*n* = 14)	Geology (*n* = 12)	Physics (*n* = 9)	Chemistry (*n* = 9)
Teaching methods						
Lecture	.21	.11	.25	.28	.20	.20
Lecture: Premade visuals (e.g., PowerPoint)	.48	.18	.69	.76	.35	.36
Lecture: Handmade visuals (e.g., chalkboard)	.44	.78	.17	.18	.55	.59
Lecture: Demonstration	.03	.00	.01	.01	.13	.04
Lecture: Interactive (i.e., 2+ questions posed)	.02	.02	.06	.01	.01	.01
Small group work	.09	.06	.16	.03	.13	.08
Desk work	.06	.12	.06	.01	.07	.04
Pedagogical strategies						
Movement	.07	.09	.07	.04	.11	.05
Humor	.07	.08	.12	.06	.03	.07
Illustration/anecdote	.10	.03	.12	.15	.17	.06
Organizational marker	.08	.07	.07	.09	.07	.11
Emphasizes topics	.05	.04	.06	.06	.03	.08
Assessments	.05	.02	.06	.04	.08	.06
Student–teacher interactions						
Rhetorical questions	.09	.13	.07	.08	.09	.08
Display questions (e.g., What is X?)	.36	.51	.38	.24	.35	.30
Comprehension questions (e.g., Do you understand?)	.11	.22	.07	.04	.11	.09
Student comprehension question	.09	.13	.11	.06	.06	.08
Student response to question	.13*	.38*	.11*	.12*	.15*	.17*
Cognitive engagement						
Articulate	.06*	.08*	.10*	.03*	.16*	.06*
Recall/memorize information	.18*	.26*	.31*	.29*	.22*	.24*
Problem solving	.12	.27	.06	.02	.16	.09
Making connections to world	.09	.03	.10	.13	.16	.05
Instructional technology						
Chalkboard	.38	.70	.14	.12	.57	.41
Overhead projector	.07	.01	.07	.15	.03	.09
PowerPoint	.47	.13	.74	.70	.37	.40
Clickers	.05	.02	.05	.05	.10	.06
Digital tablet	.07	.13	.05	.00	.00	.14
Movies or simulations	.01	.00	.02	.01	.00	.00

*These proportions only include data from two of the three institutions in the study. The version of the Teaching Dimensions Observation Protocol used to collect data at the first institution either did not contain these codes or contained different definitions of the codes from the other research sites.

these revised codes, the sample reflects two institutions ($n = 38$) only. More information about the TDOP, including the actual instrument and technical background, is available at http://tdop.wceruw.org.

Teaching practices of math and science faculty

Results from classroom observations using the TDOP provide an in-depth glimpse into the dimensions of teaching being used by the faculty in our sample. All data in Table 1 represent the proportion of times that a particular code was observed across all 2-minute intervals.

These data provide an accounting of classroom practice that demonstrates how singular measures of the teaching methods used by faculty obfuscate nuances of actual classroom behaviors. For example, the data highlight the prevalence of pedagogical strategies such as the use of illustrations and anecdotes (10% observed in all 2-minute intervals), humor (7%), and organizational markers (8%) that represent another dimension of instruction beyond the use of teaching methods. Similarly, faculty use different approaches to the asking of questions during class, such as rhetorical questions that students are not expected to answer (9%) or more open-ended questions known as display questions that solicit specific information from students (36%). Different types of possible cognitive engagements were documented, such as recalling and memorizing information (18%) and problem solving (12%), as well as the technologies used by faculty in the classroom. Finally, the frequency with which each code was observed varied considerably among disciplinary groups, thus underscoring the importance of accounting for disciplinary context.

These data also cast doubt on the use of the term *lecturing* as an adequate descriptor for any mode of instruction that involves the instructor speaking to his or her classroom. For example, although faculty do indeed speak to their classrooms for extended periods of time (21%), in many cases faculty lecture with specific types of instructional technology including PowerPoint (48%), chalkboard writing (44%), and demonstration equipment (3%). Another less frequently used variant is the Socratic lecture, in which questions are frequently posed to students and their answers used to guide the course of discussion (2%). Furthermore, these types of lecturing are used in conjunction with the other dimensions of teaching in a variety of ways, which means that two individuals who both frequently lecture may actually exhibit very different teaching behaviors.

Biology instructors who regularly lecture

The complexity underlying the simplistic term *lecturing* can be succinctly illustrated by contrasting network graphs depicting the multidimensional teaching practices of two instructors who both regularly use different types of lecturing (see Figure 1). Both instructors were observed twice while teaching introductory biology courses in a large research university setting. The lines connecting the different TDOP nodes vary in thickness on a scale of 1 to 5 depending on the frequency with which each pair of codes was coobserved in the same 2-minute interval. Thus, thicker lines represent more frequently co-observed instructional practices. The positions of the nodes are also informative. Using an iterative procedure, we moved the nodes to locations in the graph that minimize the variation in line length. Thus, each figure should be read as a "snapshot" of an individual's teaching behaviors as measured by the specific codes that he or she was more frequently observed enacting in the classroom. It is important to note that the temporal nature of teaching is obscured in these graphics, and thus they necessarily represent a simplified depiction of the complexity of classroom instruction.

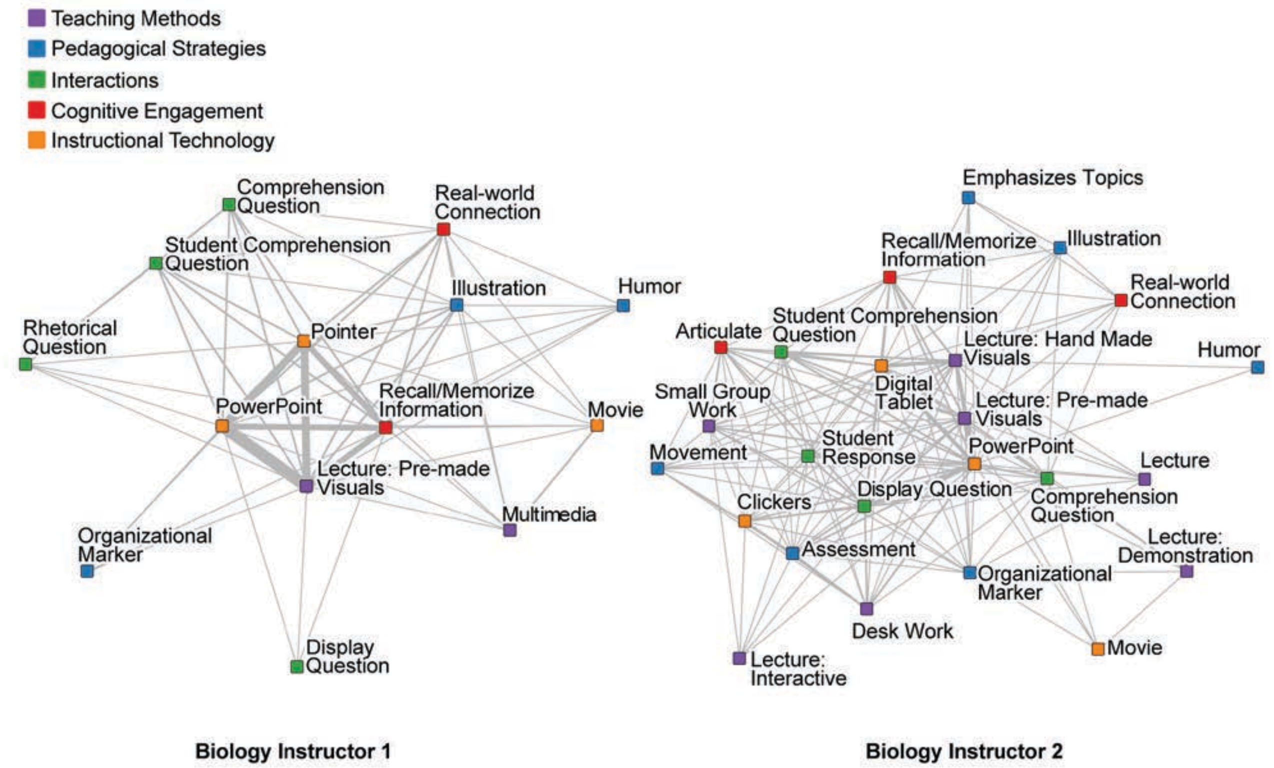
From left to right, each instructor was observed using some form of lecturing in 96% and 75% of observed 2-minute intervals, respectively. However, it would be a mistake to characterize both instructors' teaching styles as lecturing. To the contrary, Instructor 1 lectures through the use of premade PowerPoint slides and supplements this practice with, among other dimensions of instruction, illustrations and anecdotes, humor, conceptual questions, and multimedia. Meanwhile, Instructor 2 uses all five forms of lecturing and supplements this practice with a variety of cognitive engagements, student interactions, instructional technologies, and pedagogical strategies. Thus, the graphs illustrate that although both instructors lecture for the vast majority of the class, each uses a dramatically different repertoire of teaching behaviors to convey the course material to their students.

Recommendations

More detailed and multidimensional descriptions of teaching such as those obtained through the use of the TDOP instrument can be useful in four ways. Underlying each of these recommendations is the argument that descriptive research about postsecondary science teaching is critically important, much the same way that careful observations of phenomenon are central to the scientific method. Given limitations inherent in commonly used methods to study teaching, such as self-reported surveys or evaluative observations, more rigorous and objective descriptions of instructional practice as provided by protocols

FIGURE 1

Affiliation network graphs of two biology instructors teaching lower division courses.



such as the TDOP can have multiple applications, as follows:

1. *To provide rich insights of teaching practice at departmental and/or institutional levels.*

Data regarding teaching and learning are often used by institutional and departmental leaders to track student progress, to evaluate teaching, and as the basis for long-range strategic planning. However, at the present time it is uncommon for leaders to use classroom-based data for these purposes, with end-of-semester student evaluations being the most common source of information about faculty practice. Data provided from rigorous classroom observations would provide more accurate and detailed accounts of teaching that could be used to track changes in instruction over time, to evaluate the efficacy of instructional interventions, and to generally increase administrator's appreciation for the types of instruction taking place in their departments and institutions. Research on educational reform in K–12 districts and schools demonstrates that when leaders do understand and appreciate local practices, they are in a much better position to design effective interventions and policies (Spillane et al., 2001). We suggest that conceptions of instruction based on a simple dichotomy of lecturing versus interactive teaching may be unnecessarily off-putting to faculty whose awareness of their own pedagogical practice is likely more nuanced than captured by such a perspective (see Henderson & Dancy, 2008).

2. *To inform faculty professional development sessions.*

Reports about teaching are also used by campus-based professional development programs as part of one-on-one coaching and/or mentoring sessions. With an instrument like the TDOP, a detailed account of teaching can be obtained for an individual while avoiding a priori judgments about the quality of an instructor's teaching, thereby removing a potentially threatening element to professional development. When incorporated into formal professional development efforts such as new faculty orientations, these accounts can be used to spark self-reflection for individual faculty and as a way for faculty developers to gauge an individual's progress or growth over time. That said, we caution against the use of classroom-based data as a singular measure of teaching quality or efficacy. Any attempt to assess instructional quality should be based on a variety of measures and data sources, including student outcomes (see Shulman, 1987).

3. *To provide more accurate accounts of classroom teaching for policy makers.*
In seeking to improve how science and math are taught to our nation's undergraduate students, policy makers often encourage faculty to eschew lecturing in favor of interactive teaching methods. In doing so, however, policy makers are oversimplifying the problem. With more accurate descriptions of existing practice, policy makers can target scarce resources toward programs that align research-based practices with actually existing practices.

4. *To explore the relationships between classroom teaching and student learning.*
It is important that apart from the cognitive engagement codes in the TDOP, these data reveal little about how students respond to these different types of instruction, accounts of which must originate from the students themselves. Future research in this area should explore the relationships between specific types of teaching behaviors and student interpretations of the quality and efficacy of these behaviors. Insights into these complex phenomena would shed light on the types of instruction (e.g., lecturing) that students perceive as being the most beneficial for their own studying and learning. ■

References

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*, 529–542.

Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement* (CPRE Research Report No. RR-43). Philadelphia, PA: Consortium for Policy Research in Education Research, University of Philadelphia, Graduate School of Education.

DeAngelo, L., Hurtado, S., Pryor, J. H., Kelly, K. R., & Santos, J. L. (2009, February). *The American college teacher: National norms for the 2007–2008 HERI faculty survey*. Los Angeles, CA: Higher Education Research Institute, University of California–Los Angeles.

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

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

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

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,* 79–91.

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

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

Morell, T. (2004). Interactive lecture discourse for university EFL students. *English for Specific Purposes, 23*, 325–338.

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

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

President's Council of Advisors on Science and Technology. (2012, February). *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. Available at http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf

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

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

Shulman, L. S. (1987). Assessment for teaching: An initiative for the profession. *Phi Delta Kappan, 69,* 38–44.

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

Turpen, C. C., & Finkelstein, N. D. (2009). Not all interactive engagement is the same: Variations in physics professors' implementation of peer instruction. *Physical Review Special Topics—Physics Education Research, 5*(2), 1–18.

Matthew T. Hora (hora@wisc.edu) *is an assistant scientist and* ***Joseph J. Ferrare*** *is an associate researcher, both at the Wisconsin Center for Education Research, University of Wisconsin–Madison.*