

A multidimensional assessment of reformed teaching practice in geoscience classrooms

Rachel Teasdale¹, Karen Viskupic², Julie K. Bartley³, David McConnell⁴, Cathryn Manduca⁵, Monica Bruckner⁵, Dori Farthing⁶, and Ellen Iverson⁵

¹Department of Geological and Environmental Sciences, California State University, 400 W. 1st Street, Chico, California 95929-0205, USA

²Department of Geosciences, Boise State University, 1910 University Drive, Boise, Idaho 83725, USA

³Department of Geology, Gustavus Adolphus College, 800 W. College Avenue, St. Peter, Minnesota 56082, USA

⁴Department of Marine, Earth & Atmospheric Sciences, North Carolina State University, 2800 Faucette Drive, Raleigh, North Carolina 27695-8208, USA

⁵Science Education Resource Center, Carleton College, One North College Street, Northfield, Minnesota 55057, USA

⁶Department of Geological Sciences, State University of New York at Geneseo, 1 College Circle, Geneseo, New York 14454, USA

ABSTRACT

This study analyzed quantitative and qualitative data from classroom observations combined with instructor survey results to characterize the application of reformed teaching practices in undergraduate geoscience classes in the United States. Trained observers used the Reformed Teaching Observation Protocol (RTOP) to score 204 geoscience classes. Observed faculty represent a diversity of institutions, teaching rank, and years of experience. Classrooms observed included introductory and upper-level undergraduate courses that ranged in size from 6 to 275 students. Total RTOP scores do not correlate with class size, class level, institution type, instructor gender, instructor rank, or years of teaching experience. Classroom instruction was separated into three categories based on total RTOP scores: Teacher Centered (≤ 30), Transitional (31–49), or Student Centered (≥ 50). Statistical analyses of RTOP subscales and individual item scores are used to identify the instructional practices that are characteristic of each category. Instructor survey responses and qualitative classroom observations provide additional details about instructional practices common within each instructional category. Results of these analyses provide a coherent picture of instructional strategies used in geoscience classrooms. Instruction in the most Student Centered classrooms differs from that in Transitional and Teacher Centered classrooms in at least one of three ways. Student Centered classes are more likely to include (1) students engaged in class activities with one another; (2) activities in which instructors assess student learning and adjust lessons accordingly; and (3) opportunities for students to answer and pose questions that determine the focus of a lesson.

INTRODUCTION

Student-Centered Practices in the Sciences

Reformed teaching is defined as a set of practices that place priority on student agency to promote active learning, problem solving, and critical thinking (Sawada et al., 2002). Active learning occurs through diverse student-centered teaching practices, and can be described as strategies that “promote concep-

tual understanding through interactive engagement of students” (Hake, 1998, p. 65). In a classroom setting, the phrase active learning describes time during which students actively work on problems or questions (Marlowe and Page, 2005; Eddy et al., 2015) rather than passively listening to an instructor. Student-centered practices engage students in peer-to-peer interaction that often emphasizes higher order thinking (Freeman et al., 2014). The use of these strategies has been shown to increase student engagement (Umbach and Wawrzynski, 2005), to improve student mastery of course material across STEM (science, technology, engineering and mathematics) disciplines (Prince, 2004; Kuh et al., 2005; Fairweather, 2009; Singer et al., 2012; Freeman et al., 2014; Kober, 2015), and to reduce the achievement gap among student populations (Haak et al., 2011; Eddy and Hogan, 2014). Positive correlation between the presence of active learning and student mastery of course material has been documented in introductory and upper-level geoscience courses (e.g., Yuretich et al., 2001; McConnell et al., 2003; Yuretich, 2004; Kortz et al., 2008; Goldsmith, 2011; Dohaney et al., 2012) where students also report satisfaction with use of reformed teaching methods (Yuretich et al., 2001; McConnell et al., 2003; Yuretich, 2004; Dohaney et al., 2012).

Evidence from studies of student learning and student perceptions supports a positive relationship between active learning strategies and student performance and retention (Hake, 1998; McConnell et al., 2003; Epting et al., 2004; Knight and Wood, 2005; Umbach and Wawrzynski, 2005; Michael, 2006; Freeman et al., 2014). However, despite evidence in favor of adoption of reformed teaching practices in STEM, there remains a significant proportion of instructors who have not embraced student-centered strategies (e.g., Henderson and Dancy, 2007; Wieman et al., 2010; Singer et al., 2012). Given this evidence, it is of value to the geoscience community to document the use of student-centered practices in geoscience classrooms so that the community can reflect on the state of pedagogical practices in undergraduate geoscience education and identify practices that may help instructors seeking to shift their pedagogy to incorporate more student-centered strategies. The Classroom Observation Project (Science Education Resource Center, 2014) was established through the On the Cutting Edge Program to provide data about the types of teaching strategies used in undergraduate geoscience classrooms across the nation and to document the degree to which reformed teaching practices are employed in geoscience classes.

Student-Centered Teaching in the Geosciences

There are many examples of student-centered teaching practices used in individual undergraduate geosciences classrooms (e.g., McConnell et al., 2003; Greer and Heaney, 2004; Yuretich, 2004; Kortz et al., 2008; Goldsmith, 2011). While these publications provide evidence that student-centered teaching strategies are used in some geoscience classrooms, they do not provide information about how widespread the use of such teaching strategies might be among geoscience instructors nationwide.

More than 2000 instructors responded to questions about the frequency of use of various teaching methods in the On the Cutting Edge 2004 national survey (Macdonald et al., 2005), which showed that traditional lecture was the most commonly used classroom practice, with 66% of introductory course instructors and 56% of upper-level course instructors reporting using lecture in nearly every class (Macdonald et al., 2005). While the survey results of Macdonald et al. (2005) indicate use of instructional strategies other than lecture, most instructors report using such strategies infrequently. These survey results suggest that undergraduate geoscience classes use student-centered strategies to varying degrees. Now, more than a decade later, it is worth investigating the degree to which student-centered strategies are used in undergraduate geoscience classrooms, and to investigate the degree to which instructor survey results match direct observation of classroom practice, because some studies have called into question the reliability of self-report data from instructor surveys (e.g., Fung and Chow, 2002; Ebert-May et al., 2011).

Direct observation of teaching practices using an observation protocol provides another description or measure of student-centered strategies used in classrooms. Budd et al. (2013) developed a rubric to accompany the Reformed Teaching Observation Protocol (RTOP; Sawada et al., 2002) and used it with two trained observers to describe the teaching practices of 26 instructors in introductory geoscience courses at a variety of institution types. Budd et al. (2013) used RTOP observation data to categorize instruction as Teacher Centered (RTOP score ≤ 30), Transitional (RTOP score 31–49), or Student Centered (RTOP score ≥ 50), and indicated that about one-third of introductory geoscience instructors engage in significant student-centered practices in their courses. While valuable, these data describe only a few geoscience courses at the introductory level. To date, the degree to which student-centered teaching strategies are used in all levels of college geoscience classes has not been documented, nor do we know which particular strategies geoscience instructors most often use.

Research Questions

The Classroom Observation Project aims to describe the extent of student-centered teaching used in undergraduate geoscience classrooms at all instructional levels and at a variety of institution types (Science Education Resource Center, 2014). To achieve this goal, we identify two research questions. (1) How do teaching practices observed in undergraduate geoscience classrooms vary with demographic variables such as class size, class level, and

instructor teaching experience? (2) What teaching practices used in geoscience classrooms characterize and differentiate Teacher Centered, Transitional, and Student Centered learning environments?

By answering these two questions we seek to identify common variables and practices that are associated with student-centered teaching in the geosciences with the expectation that such knowledge could be used to promote student-centered teaching practices among geoscience instructors.

Ways to Characterize Teaching Practices

Given increasing attention to pedagogy in college-level STEM classrooms as a means to improving student performance, retention, and interest in STEM disciplines, diverse studies have collected and analyzed classroom data in a variety of ways. Studies that address the perspectives of instructors include investigations of teacher beliefs (Kane et al., 2002; Roehrig and Kruse, 2005), examinations of instructor training (Gibbs and Coffey, 2004; Postareff et al., 2007), interviews with instructors (e.g., Markley et al., 2009), and instructor self-report surveys (e.g., Macdonald et al., 2005; Chang et al., 2011; Wieman and Gilbert, 2014). Direct observations of classroom practices are also used and often employ formal rubrics and observation protocols to systematically measure teaching practice (e.g., Sawada et al., 2002; Hora and Ferrare, 2013; Smith et al., 2013; Lund et al., 2015). This work combines results of direct observation of classroom practices with instructor-reported survey responses. The following is a review of common observation protocols to provide context and justification for our instrument choice.

Common Observation Protocols Used in STEM Disciplines

Quantitative methods to describe classroom practices vary widely in the types of instruction that they aim to identify and in the types of data they produce. The RTOP was developed by the Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT) and provides a standardized means for detecting the degree to which classroom instruction uses student-centered, engaged learning practice (Lawson et al., 2002; MacIsaac and Falconer, 2002; Sawada et al., 2002). Use of the RTOP instrument captures many dimensions of student-centered teaching practices, including focus on a conceptual framework, group work, class discussion, hypothesis testing, and emphasis on exploration in the classroom. The RTOP also assesses student and instructor interactions, which are important for understanding student engagement (e.g., Good and Brophy, 1997; Wainwright et al., 2004). The RTOP instrument has 25 items grouped into 5 subscales (Sawada et al., 2002), and each item is scored from 0 (never occurred) to 4 (thoroughly demonstrated).

Other observation protocols that use detailed coding strategies include the Teaching Dimensions Observation Protocol (TDOP; Hora, 2015) and the Classroom Observation Protocol for Undergraduate STEM (COPUS; Smith et al., 2013). Both were developed specifically for post-secondary nonlaboratory courses. With both protocols, trained observers document student and

instructor activities during each 2 minute time interval of an observed class period (Hora and Ferrare, 2013; Smith et al., 2014) by coding for occurrences of specific student and teacher behaviors (Hora and Ferrare, 2013; Smith et al., 2013). The TDOP has 33 possible activities to code in 5 categories (Hora, 2015), which are reduced to 25 elements in the COPUS. Similarly, the Practical Observation Rubric to Assess Active Learning (PORTAAL) was developed from discipline-based education research to capture evidence of students engaged in research-based elements of active learning in observed STEM classrooms (Eddy et al., 2015).

Different observation protocols have different outputs. RTOP results are quantitative scores that permit straightforward comparisons across observations and observers. Other observation protocols, like PORTAAL, COPUS, and TDOP, produce an average frequency or duration of time spent on each coded element, useful in informing the presence, absence, or proportion of time spent on specific items, but can be more challenging to report for research purposes. Smith et al. (2013) suggested that observations using protocols that result in numerical scores, such as the RTOP, may be perceived as judgmental and thus awkward to share with the observed instructor. As such, protocols like COPUS and TDOP that report the proportion of class time during which coded activities occurred may remove perceived judgment by documenting only whether something did or did not happen (e.g., Chism, 2007). However, such protocols also remove the ability to measure (or document) the quality or depth of observed activities (e.g., clicker questions would be recorded without regard for the type of question asked or the type of thinking students were expected to do to answer the question). Such coding is difficult to distill when considering a single class period, and simultaneous coding for multiple elements can be challenging to interpret (Lund et al., 2015). In addition, 2 minute time blocks may include multiple activities (e.g., writing on the board, lecturing, and asking questions), all of which go into the reported percentages of activities during the class period, making it difficult to compare class periods among faculty or class types.

Additional criticisms of observation protocols include the training, organization, and logistical planning required. For example, training to use the COPUS can be as short as 1.5 hours (Smith et al., 2013), whereas the RTOP and TDOP have been criticized for requiring extensive observer training (e.g., Smith et al., 2013; Eddy et al., 2015). Some protocols, such as PORTAAL, require that observers record multiple types of information at frequent intervals during a class period, which led Eddy et al. (2015) to suggest that live observations using PORTAAL are impractical and should instead be made from video recordings that allow the observer to pause or review the class. The use of videos in making PORTAAL observations (e.g., Walkington and Marder, 2013; Eddy et al., 2015) requires an extra element of planning and willingness on the part of the instructor to allow video cameras in the classroom. In addition, having cameras in the classroom may raise concerns regarding student privacy.

Given the strengths and weaknesses of the array of observation protocols available, we selected the RTOP rubric for the large-scale quantitative research goal of characterizing undergraduate geoscience classrooms. The overarching

strength of the RTOP is that it was developed from discipline-based education research and uses trained observers with content knowledge to assess the quality and depth of use of research-based pedagogies (Eddy et al., 2015). Training required for RTOP observers is extensive, but interrater reliability for observers is high, and differences in RTOP scores reflect distinct instructional practices (Sawada et al., 2002; Budd et al., 2013).

Work presented here includes a large sample of geoscience faculty ($n = 204$) in both introductory and upper-level undergraduate geoscience courses to represent all stages of the undergraduate experience. This work also combines quantitative RTOP scores with analysis of qualitative observer comments and instructor survey data. These data broadly characterize the teaching strategies used in geoscience classrooms in a variety of class types at a broad range of institutions by diverse instructors across the United States.

METHODS

A team of classroom observers, trained by the On the Cutting Edge sponsored Classroom Observation Project (Science Education Resource Center, 2014), used the RTOP to complete more than 200 classroom observations in a wide range of geoscience classrooms across the United States. Total RTOP scores are used to describe the degree and variability of teaching practices used in different types of undergraduate geoscience classrooms. To characterize and differentiate teaching practices observed in Teacher Centered, Transitional, and Student Centered learning environments, we analyze RTOP scores, instructor responses to a teaching practices survey, and written comments recorded by classroom observers. This multidimensional data analysis provides quantitative and qualitative characterizations of teaching practices in a large number of geoscience classrooms.

Classroom Observations

Given the goal of describing geoscience classrooms representative of those across the United States, a team of observers was drawn from a range of institution types, geographic regions, and with different discipline specializations. Classrooms to observe were initially identified based on proximity to trained observers (first ~50 observations) and later were identified with an effort to balance demographic information such as instructor gender, rank (adjunct through full professor), institution type (associates through doctorate granting, based on the Carnegie Classification [2015]) and class size at both introductory and advanced levels.

All trained observers made classroom observations using a modified version of the RTOP rubric described by Budd et al. (2013). Each of the 25 items on the RTOP is scored on a scale from 0 to 4, resulting in possible total RTOP scores from 0 to 100. RTOP items are grouped into the following 5 subscales (Sawada et al., 2002):

1. Lesson Design and Implementation. The five items in this subscale examine the design and application of a lesson in engaging students. For example, items in this subscale look for evidence that the instructor takes students' prior knowledge into account and provides opportunities for students to work together as part of a learning community.

2. Content—Propositional Knowledge. The five items in this subscale examine the character and organization of the content presented by the instructor. For example, items in this subscale look for evidence that the instructor knows the content, presents it in a way that highlights fundamental concepts, and allows students to represent and connect abstract concepts with other disciplines or everyday life.

3. Content—Procedural Knowledge. The five items in this subscale examine what students are asked to do during a class to support their learning of the content. For example, items in this subscale look for evidence that students are working with the content through their use and interpretation of, e.g., models and graphs, through their formulation of predictions and/or hypotheses, and through reflections on their own learning.

4. Classroom Culture—Communicative Interactions. The five items in this subscale examine the types of interactions among students and the quality of those interactions in terms of how students are communicating and what they are communicating about. For example, items in this subscale look for evidence that questions and comments from students influence the focus and direction of classroom activities, and that talk among students takes place during a high proportion of class time.

5. Classroom Culture—Student-Teacher Relationships. The five items in this subscale examine types of interactions between the instructor and the students, and how those interactions promote student participation and learning. For example, items in this subscale look for evidence that the instructor encourages and values student participation, that the instructor listens to student ideas, and that the instructor is patient with students.

RTOP observers were trained in four cohorts. The first cohort (12) observed classrooms and watched recorded classroom sessions to test and adapt the RTOP rubric of Budd et al. (2013). The modified rubric was used by all observers for the 204 observations reported here. By developing a detailed rubric for the RTOP, Budd et al. (2013) established high interrater reliability between their two observers. In our larger observer pool, initial discussion and comparison of scored observations using the Budd et al. (2013) RTOP rubric led us to slightly modify the wording in the scoring rubric and/or comments associated with the scoring rubric in 15 of the 25 rubric items. These changes were made to clarify language and ensure consistent application of the RTOP instrument by observers for work reported here. For example, in the area of Procedural Knowledge, item 11 addresses the use of a variety of means to represent phenomena. A comment was added to clarify that asking students to interpret a graph, map, or diagram does count as interpreting phenomena.

Through iterative training observations, scoring discussions, and slight modifications to the Budd et al. (2013) RTOP rubric, the first cohort came to

consensus, and after ~8 observations and adjustments to the scoring criteria, observation scores of the cohort fell within a narrow range. The mean scores of this original cohort were henceforward considered the standard scores for subsequent training. Three additional cohorts of observers were trained from 2012 to 2014 using a three-stage process of scoring videos and discussing scores with a trained observer. At each stage if trainee scores were within one standard deviation of the standard, they advanced to the next stage of training, which included more videos, discussions, and score comparisons. The final stage in the training process required participants to score two final calibration videos. Using these final calibration videos, Cronbach's alpha was calculated to be 0.996 for all scores received from the pair of final calibration videos from observers [taken individually, the α for video 1 is 0.81 ($n = 22$) and for video 2 is 0.84 ($n = 24$)]. This exceeds the acceptable threshold for interrater reliability of $\alpha > 0.7$ (Multon, 2010). This measure of instrument reliability is consistent with that calculated for other studies using RTOP (e.g., Sawada et al., 2002; Budd et al., 2013). We also determined interrater reliability using a two-way mixed, single measures intraclass correlation (Hallgren, 2012), using data from 22 observers and 2 training videos. This method yields an intraclass correlation (ICC) of 0.928 for the total RTOP score and ICC > 0.5 for all subscales except propositional knowledge (ICC = 0.215). Our finding is consistent with other studies that show lower reliability for the propositional knowledge subscale than for other parts of the instrument (Budd et al., 2013).

Classroom observations were made by trained observers in person (not by video) between March 2011 and June 2014 as part of the On the Cutting Edge Classroom Observation Project (Science Education Resource Center, 2014) and resulted in 204 sets of quantitative observer data (RTOP scores). No single instructor was observed more than once. After the first year of observations in 2011–2012, observers also provided a description of activities that were used to assign the quantitative scores of each RTOP subscale, a summary of the class, examples of missed opportunities and highlights, and any other characteristics that might have made the class an anomaly. Observer comments are available for 172 of the 204 quantitative observations and vary in length and quality. Instructors who agreed to have a class observed were asked to sign a consent form (per institutional review board requirements) and to complete the RTOP instructor survey (Appendix), which is a subset of the 2009 On the Cutting Edge teaching practices survey (Manduca et al., 2011). Participating instructors provided demographic data, information about the observed class, and information about their typical teaching practices. Instructors from 203 of the 204 observed classrooms submitted survey responses.

Total RTOP scores are compared to demographic variables to look for factors that may influence the degree of student-centered teaching practices used in geoscience classrooms. Total RTOP scores are also used to assign each of the 204 observations to an instructional category according to the classification established by Budd et al. (2013): Teacher Centered (RTOP score ≤ 30), Transitional (RTOP score 31–49) or Student Centered (RTOP score ≥ 50). Following this categorization we use the following:

1. Discriminant function analysis to identify RTOP subscales and items that most strongly characterize and differentiate classes in each instructional category.

2. Comparison of survey responses with quantitative RTOP data to connect observations of single class sessions with instructor report of general teaching practices.

3. Analysis of qualitative classroom observer comments to identify specific teaching practices that characterize and differentiate classes in each instructional category.

While the three instructional category labels, Student Centered, Transitional, and Teacher Centered (Budd et al., 2013), have been criticized as being value laden (e.g., Hora, 2015), they are already in use, and therefore we use them here to avoid inventing new category titles that may cause confusion. The Transitional category is used only to indicate that the RTOP score is intermediate between the Teacher Centered and Student Centered categories (Budd et al., 2013), and does not suggest that an instructor is attempting to pass through this category to get to another. Furthermore, the instructional category labels and RTOP scores are not meant to imply that single observations represent an instructor's entire teaching practice, but they are useful in characterizing the learning environment observed during a single class period.

Discriminant Function Analysis of RTOP Scores

To explore which RTOP elements are most strongly associated with membership in a particular instructional category, we performed a discriminant function analysis (DFA). The DFA is a statistical analysis that examines the degree to which a set of independent variables predicts the outcome of a dependent, categorical variable. In this study, independent variables are the 25 items in the RTOP instrument, and the dependent variable is the instructional category (Teacher Centered, Transitional, and Student Centered). The analysis identifies the combination of RTOP items that most reliably distinguishes among instructional categories by creating functions from weighted combinations of the independent variables that maximize variance in the dependent variables. In other words, this test allows us to examine which items are most characteristic of each instructional category. The fit of the model is evaluated by determining the proportion of cases for which the discriminant functions correctly predict the resultant instructional category. The relative contributions of each independent variable to a function are described by its canonical loading.

SPSS 24 (Statistical Package for the Social Sciences software) was used to perform a series of DFAs to probe the relationship between RTOP score elements and instructional category. The DFA and statistical assumptions were tested using built-in functions of SPSS, and the results are discussed in the following section. In the initial DFA, instructional category was used as the dependent variable and the five subscales were used as independent variables (Analysis A). In a second analysis, the 25 individual RTOP rubric items were used as independent variables (Analysis B). This approach was refined in a

third analysis (Analysis C), in which a DFA was performed using only the most predictive rubric items to construct a model that predicted instructional category well (>80%) with the smallest possible number of predictor variables.

Assumptions and Limitations of DFA

Applied to these data, a DFA has only limited statistical utility, but it can be used to probe the variables that are most likely to characterize instructional category. Because the independent variables (items or subscales) compose the dependent variable (RTOP score), they are not unrelated. However, because each independent variable contributes equally to the total RTOP score, a finding that some items are more strongly predictive of instructional category might indicate that these variables deserve further investigation.

A DFA assumes that all independent variables are normally distributed, that the variables are homoscedastic (variance within dependent variables is uncorrelated to independent variables), and that participants' scores are independent of one another. RTOP items were homoscedastic, as standardized residual values were uncorrelated with standardized predicted values; furthermore, each participant was scored independently of all other participants, making each individual participant's score independent of the others. In contrast, independent variables were not always normally distributed. The non-normal character of the distribution is due to the ordinal nature of the variable (five discrete states for each variable), producing relatively flat, left-skewed or right-skewed distributions. Previous work suggests, however, that the DFA is most sensitive to outliers and can be applied if a non-normal distribution is mainly due to skewness, rather than to data outliers (Tabachnick and Fidel, 1996). The DFA also assumes that participants were sampled randomly. Although not a random sample, participants in this study were not selected systematically to represent a particular subgroup of instructors; rather, the study used a purposive sample to include a wide cross section of instructors, course levels, and institution types. The statistical power of a DFA is reduced when independent variables covary. In this study, most RTOP items correlate with other RTOP items to some degree; i.e., classrooms with high scores in one item are more likely to have high scores on other items. However, only 3 pairs of items had correlations higher than 0.5 (items 2, 18; items 6, 7; and items 16, 18), and none had correlations above 0.7. No pairs of subscales had correlations greater than 0.5. Because of these modest violations of assumptions, we do not expect the DFA to have a high level of statistical power; nevertheless, the analysis can be used as an exploratory tool to probe the items that are most predictive of instructional category.

Instructor Surveys

Instructor surveys (Appendix) were used to connect RTOP observation data from a single class with participants' self-reported typical teaching practices. We examined instructor responses regarding the proportion of time they spend on activities during typical class periods and on how frequently they use

each of several common teaching strategies: traditional lecture, lecture with demonstrations, instructor-posed questions answered by individual students, instructor-posed questions answered simultaneously by the whole class, small group discussions or think-pair-share, whole-class discussions, and in-class exercises. The frequency with which instructors use each of these strategies was reported on a scale from 1 (never) to 5 (nearly every class). Instructors were also asked to estimate the percentage of class time spent on student activities, questions, and discussion. Total RTOP scores and scores on each RTOP subscale were considered in an attempt to connect teaching practices reported in the instructor surveys with teaching practices identified during classroom observations. Responses to each survey item were compiled and binned according to instructional category (Teacher Centered, Transitional, and Student Centered).

Reliability and Limitations of Instructor Surveys

While some studies have found that instructor surveys do not reliably represent teaching practices (e.g., Ebert May et al., 2011), the use of self-reported survey data in other observation projects has found good correlation between practices reported and teaching practices observed in use by math teachers (0.85 correlation; Mayer, 1999), which suggests that survey data are reliable. A caveat, however, is that in some cases, there may be variations in the way instructors define different strategies, and our data set does not allow us to infer how respondents may have interpreted particular survey items. For example, a “whole class discussion” might suggest an instructor-facilitated discussion among a large proportion of individuals in a class, or might be interpreted as opportunities for students to provide individual or clicker responses to instructor-posed questions. Correlations between observer and instructor survey data (see following) and success in previous work suggest that observations of single class periods reflect teaching practices used by faculty in their classes (e.g., Mayer, 1999; Ebert-May et al., 2015).

Observer Comments from Classroom Observations

Qualitative analysis of observer comments from classroom observations helped identify specific attributes associated with classrooms in each instructional category. Observer comments were used to develop emergent codes by noting attributes reported by observers. An attribute was considered present only if it was specifically mentioned or described in the observer comments. These attributes were organized by theme in a codebook, independent of RTOP item and subscale definitions. Three researchers used the codebook to analyze a subset of observer comments to agree on the codebook content and to establish reliability. Based on comparison and discussion, the codebook was refined, resulting in 41 codes organized into the following six themes (Table A1 in Appendix):

- Questions: How questions are asked and answered by both the instructor and students
- Assessment: How the instructor assesses students and makes use of assessment results

- Interaction: The frequency of student-student and student-faculty interactions
- What students are asked to do: Types of activities or thinking students are asked to do
- What instructor does: Types of behaviors or lesson structures the instructor uses
- Engagement and overall mood: Types of student behavior with respect to engagement

One researcher used the codebook to analyze observer comments in the following subset of classes:

- Most Teacher Centered—Classes with the lowest total RTOP scores in the Teacher Centered instructional category; 10 classes with RTOP scores ranging from 13 to 19.
- Mean Teacher Centered—Classes with total RTOP scores around the mean (23) of the Teacher Centered instructional category (scores ≤ 30); 10 classes with RTOP scores ranging from 22 to 24.
- Mean Transitional—Classes with total RTOP scores around the median (37) of the entire data set and the mean (39) of the Transitional instructional category; 22 classes with RTOP scores ranging from 36 to 41.
- Mean Student Centered—Classes with total RTOP scores around the mean (62) of the Student Centered instructional category (scores ≥ 50); 12 classes with RTOP scores ranging from 58 to 63.
- Most Student Centered—Classes with the highest total RTOP scores in the Student Centered instructional category; 11 classes with RTOP scores ranging from 69 to 89.

This subset of 65 classes was chosen to represent the end members of the data set and the central tendency of each instructional category.

Reliability and Limitations of Observer Comments

To establish the reliability of the analysis of qualitative observer comments, 3 researchers independently coded observer comments from 10 classes with total RTOP scores ranging from 13 to 71. This sample was selected from the 65 classes included in the observer comment analysis to represent a range of RTOP scores. For each class, researchers were considered to be in agreement for an item if all three coded it the same way. Coding 41 items for 10 classes gives 410 potential agreements among the researchers. In practice, the 3 researchers had 366 agreements, or 89% agreement, indicating a high degree of reliability for the coding.

The observer comments vary in quality and were written in response to the RTOP subscales, and thus comments focus on those particular elements of classroom practice rather than representing an unscripted narrative of a class period. Many of the classes likely included attributes that were not specifically mentioned in the observer comments, and so were not coded as present. While this may seem to be a limitation of the analysis, attributes associated with instructional categories emerged indicating that there are discernable characteristics associated with each category.

■ RESULTS

RTOP Scores and Demographic Variables

The set of 204 participants includes a broad cross section of geoscience faculty. Instructor survey data indicate that observed instructors include 37% female and 63% male faculty who identify their rank as full professors (36%), associate professors (21%), lecturers (22%), assistant professors (13%), and adjunct professors (8%; Figs. 1A, 1B). Classrooms that were observed included introductory geoscience courses (58%) and upper level courses designed for majors (41%; Fig. 1C). Class sizes reported as numbers of enrolled students were binned as small (≤ 30 students, 48%), medium (31–79 students, 30%) or large (80+ students, 22%; Fig. 1D). Classrooms observed were at research or doctoral level (52%), master’s degree granting (30%), bachelor’s degree grant-

ing (7%), and associate’s degree granting (11%) institutions based on Carnegie Classification System for Institutions of Higher Education (Carnegie Classification, 2015; Fig. 1E).

RTOP scores of classrooms observed for this project range from 13 to 89 with an average score of 39.6. Our data set contains 62 observations in the Teacher Centered category (30% of observations), 92 in the Transitional category (45%), and 50 in the Student Centered category (25%). The average and range of scores for each of the five RTOP subscales are shown in Table 1. Scores for subscales 1, 4, and 5 (Lesson Design and Implementation, Communicative Interactions, and Student-Teacher Relationships) are similar; scores for subscale 2 (Propositional Knowledge) are higher, and scores for subscale 3 (Procedural Knowledge) are lower. As expected, scores on each RTOP subscale are positively correlated with total RTOP score; however, the correlation for subscale 2 is weak.

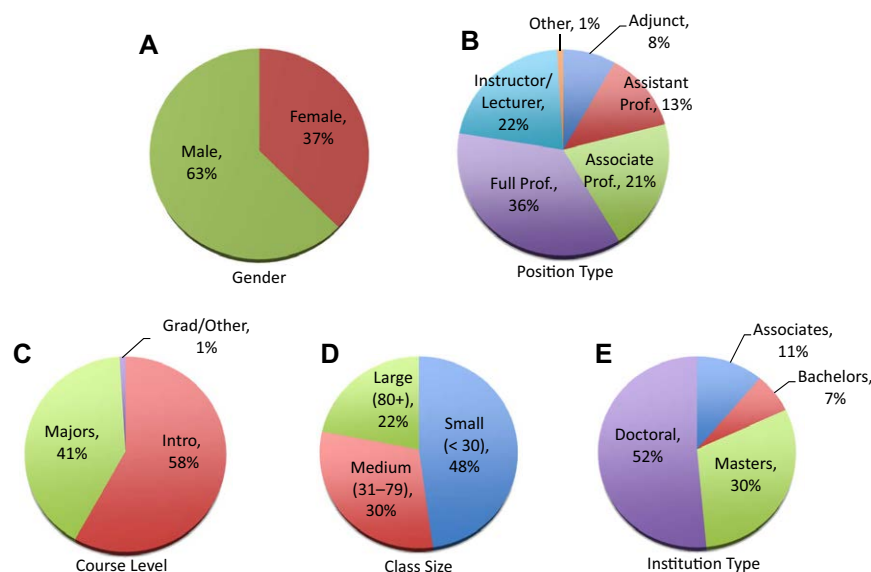


Figure 1. Institutional and demographic information for the 204 Reformed Teaching Observation Protocol (RTOP) observations in this project. Institution type corresponds to the Carnegie Classification for Institutions of Higher Education (Carnegie Classification, 2015). Course level, class size, instructor gender, and position type were provided by observed instructors in the instructor survey. Prof.—professor.

TABLE 1. AVERAGE AND RANGE OF REFORMED TEACHING OBSERVATION PROTOCOL SUBSCALE SCORES FOR THE 204 OBSERVATIONS IN THIS STUDY

RTOP Subscale	Subscale score range	Subscale score average	Correlation coefficient with total RTOP score (R ²)
1. Lesson Design and Implementation	1–18	6.4	0.8302
2. Propositional Knowledge	8–20	14.8	0.3078
3. Procedural Knowledge	0–16	4.8	0.8479
4. Classroom Culture: Communicative Interactions	0–19	6.2	0.8451
5. Classroom Culture: Student-Teacher Relationships	0–19	7.4	0.8821

Note: Total Reformed Teaching Observation Protocol (RTOP) score average is 39.6 and range of all RTOP scores = 13–89.

Total RTOP scores and the instructional category designations do not correlate with the number of years instructors have taught at the college or university level. Maximum, median, and minimum RTOP scores do not vary with any demographic factor such as institution type, course level, or instructor gender (Figs. 2A–2D). Significant statistical differences in RTOP score as a function of demographic data would require that the score ranges defined by the second and third quartiles in each category (top and bottom of boxes in Figs. 2A–2D) do not overlap with scores in other categories within the same demographic type (e.g., male or female categories within the gender demographic). Such variation is not observed in any demographic data (Figs. 2A–2D).

DFA of RTOP Scores

The results of our DFAs are shown in Figure 3 and Tables 2 and 3. Each DFA is shown as a two-dimensional plot with each function axis representing a linear combination of independent variables that maximizes variance along that axis. Such a plot identifies the independent variables that are most indicative

of membership in a particular instructional category. The more tightly clustered a category, the more alike the members of that category, in terms of the variables that make up the function axes. In the case of instructional categories defined by total RTOP score, the individual points are clustered, but the points in each category overlap with points from the adjacent category. This pattern indicates that the scores form a continuum, rather than discrete clusters; such a pattern is expected, since the instructional category divisions are artificially applied (Budd et al., 2013).

Using the five RTOP subscales as independent variables (Analysis A), the DFA produces a model that predicts total RTOP score very well (94.6% of cross-validated cases correctly identified; Fig. 3A), as expected because the subscales are combined to produce a total RTOP score. Despite the fact that each subscale is weighted equally in computing the total RTOP score, the Classroom Culture subscales (Student-Teacher relationships and Communicative Interactions) are most predictive of RTOP score: they have the highest canonical loading (correlations with Function 1 of 0.743 and 0.719, respectively; Table 2), indicating that these two subscales correlate most strongly with the discriminant functions. In addition, these variables also have high canonical discriminant function coefficients (Table 2), indicating that they contribute

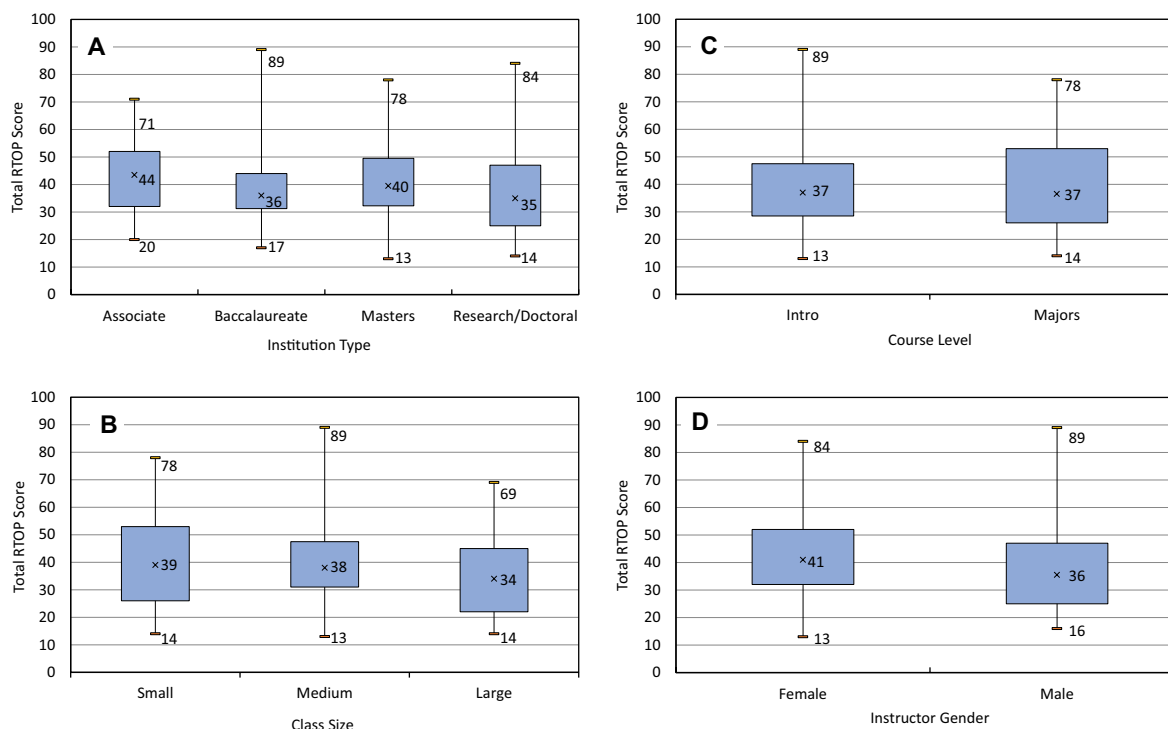


Figure 2. Relationships between total Reformed Teaching Observation Protocol (RTOP) score and institutional or demographic characteristics. For each plot, x represents the median score for that category. The blue box encompasses the second and third quartile about the median, and the thin lines represent the minimum and maximum scores recorded for that category. Categories are considered distinguishable if the 2-quartile boxes do not overlap. (A) Institution type. (B) Class size. (C) Course level and audience. (D) Instructor gender.

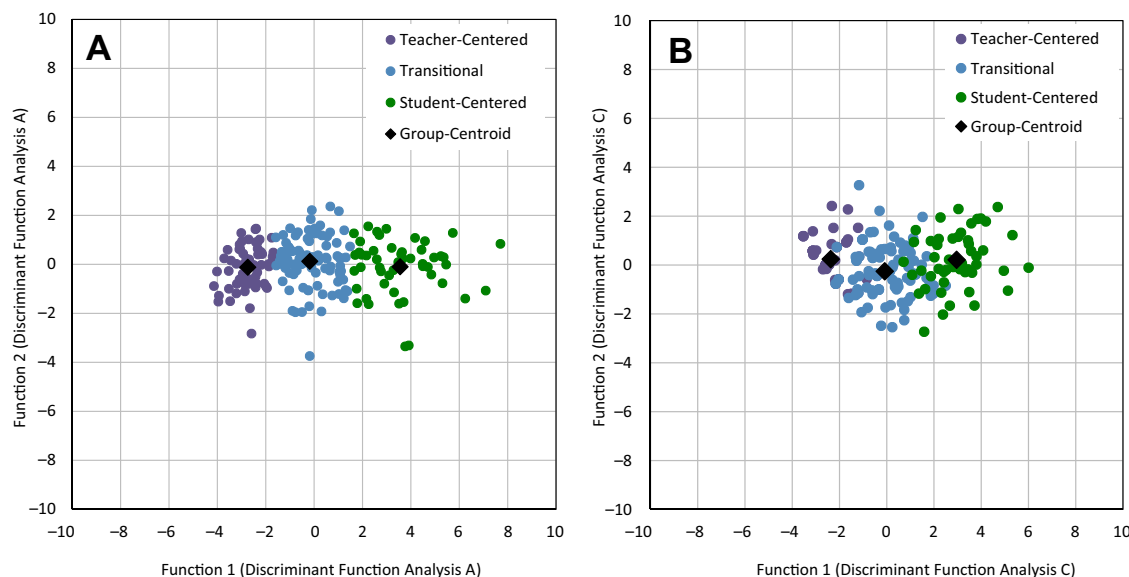


Figure 3. Results of the discriminant function analyses to explore the degree to which Reformed Teaching Observation Protocol (RTOP) subscales and individual items predict classroom category (Teacher Centered, Transitional, Student Centered). (A) Results of Analysis A evaluating the five RTOP subscales. Function 1 describes 99.7% of the variance and is significant at the 0.01 level; Function 2 is not significant. This model accurately predicts group membership in 94.6% of cross-validated cases. (B) Results of Analysis C, a discriminant function analysis using the 7 RTOP score items that had the highest correlations with Function 1 on Analysis B (Table 3), which included all 25 RTOP score items. In Analysis C, Function 1 describes 98.5% of the variance and is significant at the 0.01 level; Function 2 is not significant. This model accurately predicts group membership in 80.4% of cross-validated cases.

strongly to the function. The other three subscales have lower coefficients and lower correlations, and thus are less strongly predictive of instructional category. Propositional knowledge is the subscale with the lowest correlation with instructional category (correlation 0.242).

Using individual RTOP items in the discriminant function analysis increases the model complexity but also allows assessment of particular classroom strategies that are likely to characterize high-scoring classrooms. When all 25 RTOP items are included in the analysis (Analysis B), the model correctly classifies 90.2% of cross-validated cases (Table 3). As with Analysis A, only Function 1 is significant at the 0.01 level. Seven RTOP items have correlations with Function 1 above 0.4 (Table 3), showing the strongest predictive relationship with instructional category:

- Item 2: The lesson was designed to engage students as members of a learning community.
- Item 13: Students were actively engaged in a thought-provoking activity that often involved the critical assessment of procedures.
- Item 16: Students were involved in the communication of their ideas to others using a variety of means and media.
- Item 18: There was a high proportion of student talk and a significant amount of it occurred between and among students.
- Item 19: Student questions and comments often determined the focus and direction of classroom discourse.
- Item 24: The teacher acted as a resource person, working to support and enhance student investigations.

TABLE 2. STANDARDIZED CANONICAL DISCRIMINANT FUNCTION COEFFICIENTS AND POOLED WITHIN-GROUPS CORRELATIONS FOR A DISCRIMINANT FUNCTION ANALYSIS USING THE FIVE RTOP (REFORMED TEACHING OBSERVATION PROTOCOL) SUBSCALES AS INDEPENDENT VARIABLES (ANALYSIS A)

	Coefficient		Correlation	
	Function 1	Function 2	Function 1	Function 2
Student-Student Interaction	0.440	-0.360	0.719	-0.474
Student-Faculty Interaction	0.328	-0.368	0.743	-0.342
Procedural Knowledge	0.296	0.213	0.684	0.136
Lesson Design	0.258	0.318	0.640	0.100
Propositional Knowledge	0.301	0.759	0.242	0.846

Note: In this analysis, only Function 1 was significant ($p < 0.05$).

TABLE 3. STANDARDIZED CANONICAL DISCRIMINANT FUNCTION COEFFICIENTS AND POOLED WITHIN-GROUPS CORRELATIONS FOR A DISCRIMINANT FUNCTION ANALYSIS USING INDIVIDUAL RTOP (REFORMED TEACHING OBSERVATION PROTOCOL) ITEMS

Item	Analysis B				Analysis C			
	Coefficient		Correlation		Coefficient		Correlation	
	FN 1	FN 2	FN 1	FN 2	FN 1	FN 2	FN 1	FN 2
Q1	0.280	0.037	0.279	0.049				
Q2	0.190	0.344	0.44	0.099	0.273	-0.409	0.571	-0.061
Q3	0.075	-0.092	0.226	-0.085				
Q4	0.179	-0.353	0.325	-0.223				
Q5	-0.193	0.372	0.318	0.222				
Q6	0.126	0.321	0.124	0.234				
Q7	0.058	-0.203	0.117	0.120				
Q8	0.033	-0.106	0.082	0.011				
Q9	0.058	0.243	0.153	0.303				
Q10	0.214	-0.128	0.166	-0.085				
Q11	0.156	0.305	0.364	0.262				
Q12	0.166	0.307	0.308	0.169				
Q13	0.144	-0.307	0.432	-0.214	0.301	0.474	0.557	0.541
Q14	0.021	-0.010	0.166	-0.029				
Q15	0.177	-0.142	0.359	-0.190				
Q16	0.097	-0.091	0.420	-0.049	0.249	-0.029	0.543	0.220
Q17	0.071	0.233	0.299	-0.127				
Q18	0.128	-0.290	0.442	-0.189	-0.064	0.817	0.572	0.496
Q19	0.240	0.221	0.504	0.146	0.284	-0.496	0.654	-0.132
Q20	0.088	-0.163	0.381	-0.136				
Q21	-0.161	-0.101	0.295	-0.004				
Q22	0.104	-0.310	0.380	-0.248				
Q23	0.127	-0.515	0.391	-0.275				
Q24	0.147	0.059	0.489	-0.065	0.266	0.131	0.633	0.271
Q25	0.310	0.422	0.425	0.326	0.405	-0.436	0.552	-0.504

Note: Analysis B includes all 25 items; Analysis C includes the seven items with correlations >0.4 in analysis A. In both analyses, only Function (FN) 1 was significant (p < 0.01).

- Item 25: The metaphor “teacher as listener” was very characteristic of this classroom.

If the model is constructed to include only these 7 rubric items (Analysis C), it classifies 80.4% of cross-validated cases correctly (Fig. 3B; Table 3).

Instructor Surveys

We compared the general pedagogical practices reported by faculty to the instructional category assigned by total RTOP scores to connect what observers saw during a single class visit with what faculty report they are doing in their classrooms. In general, we found that faculty in the Student Centered instructional category reported using strategies that encourage students to work on activities in which they interact with the course content, with each other, and with the instructor; faculty in Transitional and Teacher Centered classrooms reported using such classroom activities less frequently.

Nearly 75% of instructors in Student Centered classrooms reported using small group discussions or think-pair-share activities once per week or more (survey question 9e; Appendix). In contrast, only 10% of faculty in Teacher Centered classrooms reported using small group discussions weekly and more than 50% indicated that they never use such activities (Fig. 4A). Of instructors in Transitional classes, 42% reported using small group discussion or think-pair-share activities weekly or more frequently.

Similarly, ~62% of faculty who teach Student Centered classes reported using class exercises weekly or during nearly every class period, and only 4% of faculty in Student Centered classes reported never using in-class exercises. In contrast, ~35% of faculty of Transitional and 10% of faculty of Teacher Centered classes reported using in-class exercises at least weekly.

Instructors of Student Centered classrooms also reported using other student-centered practices more frequently compared to instructors of Transitional or Teacher Centered classes. For example, instructors in all instructional categories reported using questions to individual students (Fig. 4A)

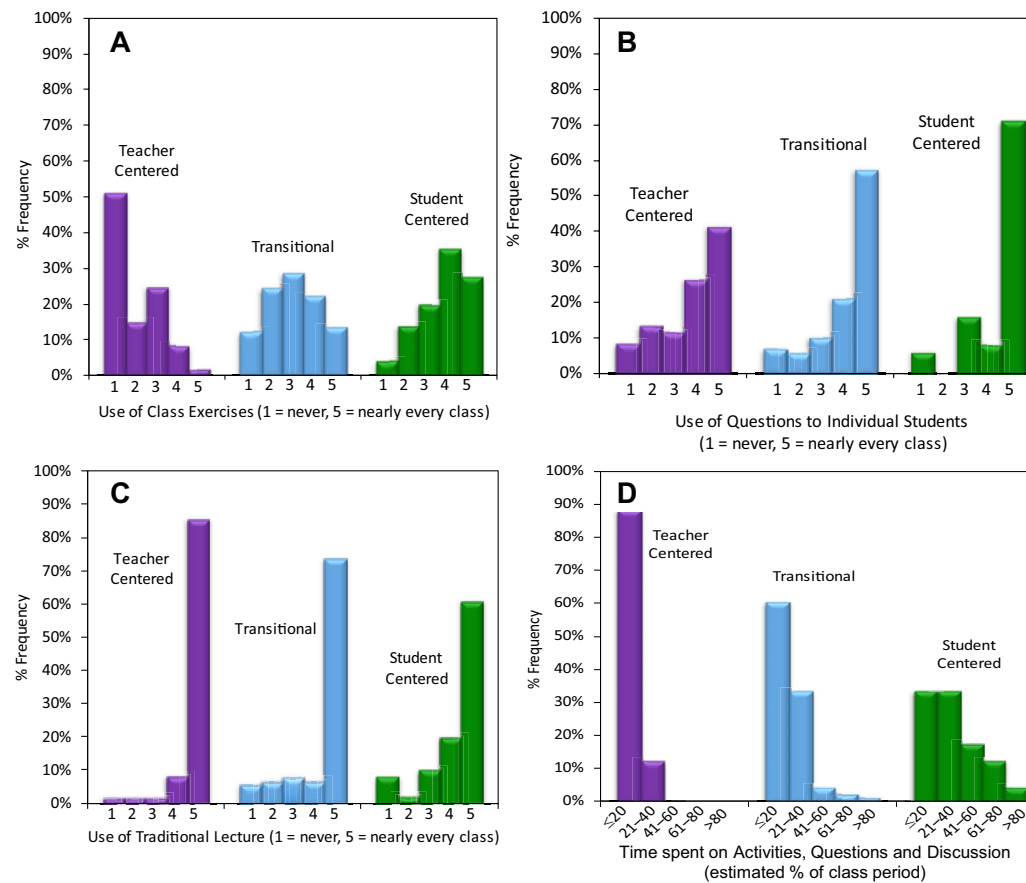


Figure 4. Instructor survey data reporting the use of specific teaching strategies, separated by instructional category. Bins for each histogram represent instructors' estimation of whether they use each strategy never (1); once or twice per term (2); several times per term (3); weekly (4); or nearly every class (5). (A) Small group work use, such as think-pair-share (TPS) activities. (B) Use of questions to individual students. (C) Traditional lecture use. (D) Estimated proportion of class time spent on activities, questions, and discussion.

more frequently than questions posed to the entire class, but there are distinct increases in the reported use of questions on a daily basis from Teacher Centered classes (41% to individuals and 30% to the whole class), to Transitional classes (57% to individuals and 45% to the whole class), to Student Centered classes (71% to individual students and 47% to the entire class).

Instructors in all classroom categories reported using traditional lecture every class period (Fig. 4B). Notably, 61% of instructors in Student Centered classes, 74% in Transitional classes, and 85% in Teacher Centered classes reported using traditional lectures every day. Only 8% of instructors in Student Centered classes reported never using lecture (Fig. 4C).

The instructor survey also asked faculty to estimate the proportion of class time spent on activities, questions, and discussion, expressed as a percentage of a typical class. Approximately 40% of faculty in Student Centered classes reported spending more than 40% of their class time on activities, questions, and

discussion (Fig. 4D), while only 4% of Transitional classes and 0% of Teacher Centered classes engaged in activities, questions, and discussion more than 40% of the time. In contrast, 64% of Teacher Centered classes reported using 20% or less of class time on activities, questions, and discussion, while 60% of Transitional and 30% of Student Centered classes spent less than 20% of class time on activities, questions, and discussion.

Observer Comments from Classroom Observations

The common codes identified for each instructional category are summarized in Table 4. There is little overlap in the most common attributes mentioned in observer comments for Teacher Centered, Transitional, and Student Centered classrooms. Even within the Teacher Centered and the Student Centered instructional categories there are differences in attributes mentioned for

TABLE 4. COMMON CLASS ATTRIBUTES AS DETERMINED BY QUALITATIVE ANALYSIS OF OBSERVER COMMENTS

	Most Teacher Centered n = 10 (%)	Mean Teacher Centered n = 10 (%)	Mean Transitional n = 22 (%)	Mean Student Centered n = 12 (%)	Most Student Centered n = 11 (%)
No or few questions asked by instructor	50	27	0	0	0
No or few questions from students	60	36	9	0	0
Students are passive or not asked to do anything	70	36	4	0	0
Lecture or traditional teaching style	50	36	32	8	0
No student-student interaction or conversation	70	80	32	0	0
Shout-out questions asked	50	50	68	8	18
Instructor asks questions of students	20	60	72	42	18
Students ask questions or volunteer ideas	20	60	32	50	18
Short wait time, first shout-out taken, or instructor answers own questions	30	0	45	0	9
Instructor reminds students of previous topics or prior knowledge	30	18	45	33	27
Instructor doesn't make use of student input or no shift in lesson	10	36	45	0	9
Students make hypotheses or predictions	0	9	45	25	36
Connections made with real world or other disciplines	40	36	55	50	27
Student-student interactions or group work	0	9	59	100	91
Lots of student-faculty interaction	0	0	14	58	36
Discussion	0	0	32	67	91
Students read graphs, maps, use data	20	27	27	67	45
Instructor circulates	0	0	9	50	45
Lesson adjustments based on student work or prior knowledge	0	0	9	33	54
Students answer open-ended questions	0	0	4	17	45
Instructor assesses students (new or prior knowledge)	10	18	18	33	45
Instructor acts as facilitator	0	0	4	8	45
No or little lecture or all or most of class is activity	0	0	4	8	45

Note: Shading indicates that the attribute is commonly mentioned in observer comments for that subset of classes, with the percentage of classrooms for which it is mentioned listed in each box. The number of classes and range of total RTOP (Reformed Teaching Observation Protocol) scores in each category are: Most Teacher Centered, n = 10, RTOP scores = 13–19; Mean Teacher Centered, n = 10, RTOP scores = 22–24; Mean Transitional, n = 22, RTOP scores = 36–41; Mean Student Centered, n = 12, RTOP scores = 58–63; Most Student Centered, n = 11, RTOP scores = 69–89.

classes with total RTOP scores near the mean for these categories and classes with total RTOP scores near the low and high extremes, respectively.

In the most Teacher Centered classrooms (total RTOP score = 13–19), observers most frequently noted that students were passive (e.g., “Students took notes; were not asked to participate”) and did not interact with each other (e.g., “No opportunity for student interaction”). This matches the common description of these classrooms as being dominated by instructor lecture or a traditional teaching style. For example, an observer in a classroom with an RTOP score of 18 noted, “This was almost a straight lecture with little opportunity for students to do anything other than observe and/or take notes.” Also commonly mentioned for the most traditional classrooms was a lack of questions from either the instructor or the students (e.g., “No questions from instructor so not really opportunities to listen to students,” and “No questions voiced by students at all”). When instructors in the most Teacher Centered classrooms ask questions, observers noted they were “shout-out” type questions, that is,

questions to which one or two students provide the single answer required before the instructor continues the lecture.

In the mean Teacher Centered classrooms (total RTOP score = 22–24), observers most frequently noted that students did not interact with each other (e.g., “Students did not interact with one another”). The principal differences between the mean and most Teacher Centered classrooms are that the mean Teacher Centered classroom instructors are more likely to ask questions, although those questions may only elicit “shout-out” responses, and that students in mean Teacher Centered classrooms are more likely to ask questions or volunteer ideas (Table 4). For example, an observer in a classroom with an RTOP score of 22 noted, “Instructor encouraged students to ask questions but didn’t provide opportunities that required their input or active participation.” Another observer in a classroom with an RTOP score of 24 noted, “One question from instructor was how to interpret a specific thing on a graph. One student answered with shout-out, instructor confirmed, elaborated.”

In the mean Transitional classrooms (total RTOP score = 36–41), observers noted that the instructors frequently ask questions of the students, but 45% of observers also commented that the wait time following an instructor-posed question is often too short to provide an opportunity for students to consider their responses (Table 4). For example, an observer in a classroom with an RTOP score of 39 noted, “Instructor has lots of questions for students, but just for the right answer as shout-outs; students participate in shout-out answers but never the chance for them to discuss together or discuss further.” Another observer in a classroom with an RTOP score of 40 noted, “Student questions were encouraged and they often resulted in the instructor clarifying ideas... There was a bit of wait time after questions, but instructor generally took the first ‘shout out.’” Other commonly mentioned attributes of mean Transitional classrooms are that students interact with each other and make predictions or hypotheses, the instructor makes connections between the lesson content and the real world or other disciplines, and the instructor reminds students of previous class topics or prior knowledge but does not make use of student input or shift the lesson in response to student input (Table 4). For example, an observer in a classroom with an RTOP score of 36 noted, “The instructor reviewed what was covered in the last class before moving on to new material, but did not assess student knowledge.”

In the mean Student Centered classrooms (total RTOP score = 58–63), observers most frequently noted that students interact with each other (e.g., “Students communicated in small groups and by writing their notes on giant post-it notes for the class”), are engaged in discussion (e.g., “Students met in small groups and discussed how they could accomplish the goal”), ask questions or volunteer ideas, and interact with data by reading graphs, maps, etc. (Table 4). Instructors in mean Student Centered classrooms often make connections between lesson content and the real world or other disciplines, interact extensively with students, and circulate throughout the classroom. For example, an observer in a classroom with an RTOP score of 59 noted, “The professor visited every group during the exercise and was mostly asking questions to explore students’ understanding of what they were doing and where they were not thinking correctly.”

In the most Student Centered classrooms (total RTOP score = 69–89), observers most frequently noted some of the same attributes found in the mean Student Centered classrooms, such as students interacting with each other, engaging in discussion and interacting with data by reading graphs, maps, etc. (Table 4). For example, an observer in a classroom with an RTOP score of 69 noted, “Students interacted with each other in small groups, during whole-class discussion and as single students at the board interacting with the instructor and class.” In the same class, the observer noted, “Students used equations and two types of plots to represent and interpret the fundamentals of the concept. They made hypotheses about mid ocean ridge processes and in some cases devised means for testing them.” Observers noted that instructors in the most Student Centered classrooms also commonly circulate throughout the classroom (e.g., “Instructor circulated around room during the activity and acted as a resource, but also to keep students on task”). In addition, in the most

Student Centered classrooms, comments indicate that lessons are commonly adjusted based on student work or prior knowledge (e.g., “Student input and discussion were encouraged and shifted the direction of the lesson”). Students in the most Student Centered classrooms are also more likely than students in mean Student Centered classrooms to answer open-ended questions, their instructors more commonly assess student knowledge, and are more likely to use little or no lecture.

In addition to identifying the most commonly mentioned attributes of each instructional category, this analysis reveals patterns in the frequency of attributes across categories. For example, instructors in all instructional categories are noted to assess student knowledge, but the percentage of classrooms in which this attribute is noted increases from the most Teacher Centered (10%) to the most Student Centered classrooms (45%). Likewise, lecture or traditional teaching style is noted in all classroom categories except for the most Student Centered; however, observers noted this attribute with decreasing frequency from the most Teacher Centered (50%) to the mean Student Centered classrooms (8%; Table 4). The percentage of classrooms in which discussion is noted by observers increases across instructional categories, from 0% in Teacher Centered classrooms to 91% in the Most Student Centered classrooms. Likewise, lesson adjustments based on student work or prior knowledge increase across instructional categories, from 0% in Teacher Centered classrooms to 54% in the Most Student Centered classrooms.

DISCUSSION

The first research question asks whether demographic variables correlate with teaching practices. Based on demographic data of observed instructors (Fig. 1), all demographic subsets such as gender, instructor rank, class size, type of institution (e.g., research or doctoral, master’s, undergraduate, or associate’s degree granting), and course type (introductory or upper level) are represented in subequal proportions. The number of classroom observations made as part of this study is significantly larger than any other similar observational survey of teaching in the geosciences we know of (e.g., Markley et al., 2009; Budd et al., 2013), and the distribution of demographic variables associated with the classroom observations establishes that we have observed a sufficiently diverse array of classes to establish whether RTOP scores, and particularly the instructional category assignments, are driven by membership in any particular demographic category (e.g., correlation between small class size and high RTOP score).

Total RTOP scores and the related instructional categories do not correlate with instructor experience, gender, class size, type of institution, or course type (Figs. 2A–2D). In short, the lack of correlation of RTOP scores with the various demographic factors of the instructors and classes we observed indicates that the demographics of the instructors, classes, or institutions are not the principal factors that determine classroom practice. Both the size of the data set and the lack of correlation between RTOP scores and demographic factors make it

appropriate to treat the entire data set in aggregate; therefore, generalizations and correlations can be applied to the population as a whole.

Characterizing the teaching practices that distinguish among Teacher Centered, Transitional, and Student Centered classrooms is the focus of our second research question. For example, items 2, 13, 16, and 18 are among the 7 rubric items that DFA analyses predict to be important in distinguishing classroom instructional categories; this is consistent with results of the instructor survey and analysis of observer comments. Scored classroom observations and faculty survey data both reported more frequent use of activities such as think-pair-share, questions, and discussions during classes in Student Centered classrooms than in Transitional or Teacher Centered classrooms (Fig. 4D; rubric item 18 in Fig. 5). Observation scores of RTOP item 18 and survey data both indicate fewer student interactions in Teacher Centered classes than in Student Centered classes, indicating concordance between these data sets. In addition, observer comments also support the quantitative RTOP scores and survey data in areas of student interactions. For example, observer comments describe students as passive more frequently in Teacher Centered classrooms and rarely or never describe students as passive in Student Centered classrooms (Table 4). Observer comments are also consistent with scores for RTOP item 18, which record "No student-student talk" (score = 0) for 94% of Teacher Centered classrooms and only for 4% of Student Centered classrooms (Fig. 5). Observers also comment on student-student interactions for most or all Student Centered classrooms and rarely or never comment on such interactions in Teacher Centered classrooms (Table 4). Thus, we find internal consistency among our analysis of all three data sets, the RTOP scores (individual items,

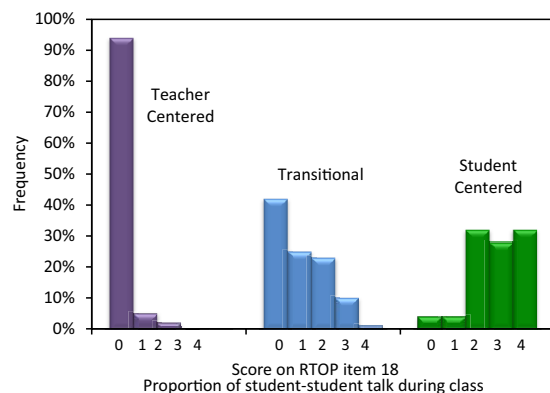


Figure 5. Direct observation data indicating the amount of student-student talk, separated by instructional category. Bins refer to the score on item 18 of the Reformed Teaching Observation Protocol (RTOP) rubric, where a score of 0 = no student-student talk; 1 = students talk to each other at least once; 2 = student-student talk occurs at least 10% of class period; 3 = student-student talk occurs more than 25% of class period; 4 = student-student talk occurs >50% of class period.

subscales, and total scores), instructor survey data, and observer comments. From all three data sets, general characteristics emerge for each instructional category based on correlations of RTOP observation scores, instructor survey data, and observer comments.

The DFA identified seven RTOP items with the strongest predictive relationship with instructional category (Table 3), indicating that there is no single strategy that can predict reformed teaching and that the diversity of teaching practices comes from a variety of reform areas (Sawada et al., 2002). Indications of the multidimensionality of reformed teaching are consistent with previous work, specifically in introductory geoscience classrooms, that suggests that there is no single pathway to reformed teaching and that a student-centered classroom environment is achieved through a holistic approach to constructivist teaching to support active engagement in the development of student knowledge (Budd et al., 2013; Hora, 2015).

The agreement among our data sets is consistent with other research that has found good correlation between self-reported data and observation data. For example, in a study of postdoctoral fellows participating in a professional development program, researchers found good correlation between RTOP observation scores and self-report data regarding participants teaching beliefs and use of learner-centered teaching strategies (Ebert-May et al., 2015). Additional work that compares surveys with observations found good correlation between survey responses regarding use of in-class activities and COPUS observation data (Smith et al., 2014). Examination of poor correlations between self-surveys and observations reported by Ebert-May et al. (2011) suggests that because surveys were taken following professional development and were administered by professional development providers, respondents may have unintentionally inflated their responses to match their intended use of professional development strategies, which later were not observed (Smith et al., 2014). Survey data are more likely to be reliable when responses are not affected by social factors or when respondents are not affected by the outcome of their responses; for example, responses might be less reliable if they are tied to promotion or considered by administrators (Desimone et al., 2010).

Based on the three internally consistent data sets reported here, the teaching practices of the three Instructional Categories can be characterized. A key feature of Student Centered classrooms is using activities that engage students in talking with each other for a high proportion of class time. This is supported by observer comments that never mention that students are passive in the most Student Centered classes. Most instructors in Student Centered classrooms also estimate that their students spend significant class time engaged in activities, questions, and discussion (Fig. 4D). This is not to say that all Student Centered classrooms spend all their time in activities; most instructors of Student Centered classes indicate that they use traditional lecture each day (Fig. 4C). It is clear that lecture remains an important instructional strategy, but as indicated by the DFA that no single RTOP rubric element or associated classroom strategy is a definitive predictor of a Student Centered classroom, lectures are used along with diverse teaching strategies (Figs. 4A–4D).

Characteristics of Transitional classrooms are similar to those in Student Centered classrooms, but the use of activities in Transitional classrooms is typically less frequent and for shorter duration. For example, faculty in Transitional classrooms reported using a lower frequency and duration of activities that engage students in talking with each other than reported by faculty of Student Centered classrooms (Fig. 4D). Observer comments corroborate the higher frequency of student interactions in Student Centered classrooms than in Transitional classes (Table 4). Similarly, more faculty in Transitional classrooms than in Student Centered classrooms reported daily use of traditional lecturing, which corresponds with observer comments that more frequently mention traditional lecturing in mean Transitional classrooms than in Student Centered classrooms (Table 4).

Teacher Centered classrooms are characterized by teaching strategies that are generally less interactive, such as more common use of traditional lecturing, which is reported by most Teacher Centered faculty as occurring every day (Fig. 4C), and is more frequently mentioned in observer comments for Teacher Centered classrooms than for Transitional classrooms (Table 4). RTOP scores indicate there are no opportunities for students to talk together about course material in 94% of Teacher Centered classrooms (item 18; Fig. 5), which corresponds to observer comments that mention no student-student interactions in 70% and 80% of most and mean Teacher Centered classrooms, respectively.

While observational and survey data indicate that geoscience instructors use a variety of pedagogies and that there is no single way to use active learning strategies in a Student Centered classroom, the consistency of results from the DFA analysis of RTOP scores, the observer comments from classroom observations, and the teaching practices reported by instructors can be used to infer which teaching strategies are most likely to shift teaching practice from Teacher Centered to Student Centered. In the following paragraphs we describe three particular characteristics that best differentiate among the instructional categories. Instructors interested in shifting their pedagogy to include more student-centered practices may wish to start by thinking about these characteristics with respect to their own classrooms. We are not suggesting that these are the only characteristics that are important, or that they are the only characteristics instructors should consider; rather, we posit that because the strategies described in the following are being used by geoscience instructors in a variety of class types, and because instructors know they are using these strategies, and therefore are actively choosing to do so, use of these strategies may be practical for other instructors as well.

Student-Student Interactions

Active participation by students in a community is integral to their ability to construct their knowledge, making learning a social practice (e.g., Bransford et al., 2000; Wenger, 2000). The seven RTOP items that the DFA indicates are most predictive of total RTOP score include three items from subscale 4 (16, 18, 19), which examine student-student interactions. Observer comments for

100% of mean scoring Student Centered classrooms (and 91% of the highest scoring Student Centered classrooms) include notes that students are working together (Table 4). In contrast, student group work is not noted for any of the 10 lowest scoring (most) Teacher Centered classrooms, and is noted in only 9% of mean Teacher Centered classrooms and 59% of mean Transitional classrooms.

The amount of time students spend talking to each other about lesson content is quantified by item 18 of the RTOP rubric, ranging from no student-student talk (score = 0), to students talking to each other at least once (score = 1), to students talking at least 10% of the class period (score = 2) to more than 50% of the class period (score = 4). Student Centered classes most commonly have students talking with each other for more than 50% of the class period (32% of classes), and 70% of Student Centered classes spend at least 25% of class time in student conversations. In contrast, 94% of Teacher Centered classrooms and 42% of Transitional classrooms include no student talk (Fig. 5).

RTOP item 2 also addresses student interactions in terms of the lesson having been designed to engage students as members of a learning community. Faculty survey data indicate that small group work and think-pair-share activities occur more frequently in Student Centered classes than in Transitional or Teacher Centered classes (Fig. 4A). The quality of student group work is measured by RTOP item 13, which examines whether students are engaged in thought-provoking activities, also one of the 7 DFA predictors for total RTOP score. Social learning experiences such as student conversations and activities are known to promote increased conceptual understanding (e.g., Stage et al., 1998; Freeman et al., 2014), making it an important element of reformed teaching, as typified in Student Centered classes.

Student-Instructor Interactions

Three of the seven RTOP items that are good predictors of total RTOP score measure the extent to which the instructor encourages student participation in determining the direction of the class period (item 19), and the extent to which the instructor supports student investigations and listens to students (items 24, and 25, respectively; Table 2). Student-instructor engagement is frequently noted in observer comments, including that the instructor acts as a facilitator (e.g., asks and answers questions as students work through an activity) and circulates throughout the room, which was noted most frequently in the mean and most Student Centered classrooms. Observers also noted that instructors in Student Centered classrooms use student work to adjust the lesson, but no such comments are reported for the mean and most Teacher Centered classrooms. Instructors in Teacher Centered classrooms are also not noted to circulate or act as facilitators of student work and learning (Table 4). Similarly, observers frequently indicate that faculty in Student Centered classrooms assess student knowledge during class, but this happens much less frequently in Transitional and Teacher Centered classrooms (Table 4). In addition to students working together to develop communities of practice to enhance learning (e.g., Wenger, 2000), it is important that faculty work with students as they construct

their knowledge (e.g., Stage et al., 1998). Faculty interaction with students has also been noted as critical to student engagement, persistence, and learning (e.g., Umbach and Wawrzynski, 2005). Thus, instructors of Student Centered classrooms act as facilitators of student learning using strategies such as circulating through the classroom as they encourage student discussions, asking leading questions, helping student groups to stay on task, and monitoring constructive learning communities.

Use of Questions

While all 7 RTOP items identified by the DFA are required for strongest predictive value, item 19 (Student questions and comments often determined the focus and direction of classroom discourse) has the highest correlation (Analysis C; Table 3). Faculty in Student Centered classrooms report that they ask questions of individual students nearly every day (71%; Fig. 4B) and are the most likely group to pose questions to the entire class on a daily basis (47%). Similarly, observers note that instructors in Student Centered classrooms asked questions frequently, and often note the use of open-ended questions (Table 4). Faculty in Student Centered classes are also most likely to use significant wait time after asking a question, allowing students ample time to think about and formulate a response. Significant wait time (e.g., 3 seconds or more) is shown to improve the quality and quantity of student responses (e.g., Stahl, 1994). In contrast, faculty survey data and observer comments note the least frequent use of questions in Teacher Centered classrooms (Fig. 4C; Table 4). Thus, each of the data sets converge on similar characteristics that can be used to describe the characteristic use of questions in Student Centered, Transitional, and Teacher Centered instructional categories.

CONCLUSIONS

Analysis of the 3 sets of qualitative and quantitative data associated with 204 observations of faculty in United States geoscience classrooms is used to address the Classroom Observation Project's research goals of describing the use and characteristics of student-centered pedagogical practices in undergraduate geoscience classrooms. A wide range of RTOP scores represent teaching practice observed, but scores do not correlate with demographic variables. The lack of correlation between RTOP scores and factors such as class size, class type (introductory versus upper level), and institution type indicates that the degree to which instructors incorporate student-centered teaching strategies, as measured by the RTOP, is dependent on instructors' pedagogical choices and not on external variables.

Further analysis of the three data sets reveals consistent patterns that identify teaching practices used in geoscience classrooms that characterize and differentiate Teacher Centered, Transitional, and Student Centered classrooms. DFA identifies seven RTOP rubric items that are most often associated with Student Centered classrooms, suggesting that those classrooms share some

broad similarities in instructional strategy. Observer comments and instructor survey data support this notion, and indicate that specific strategies that are employed frequently in Student Centered classrooms are less common in Teacher Centered and Transitional classrooms. Given that seven items are required to distinguish classroom types (rather than only one or two, for example), this work indicates that classrooms across the range of geoscience classroom types utilize a diversity of student-centered teaching practices, as indicated by previous results for a smaller population of introductory geoscience classrooms (Budd et al., 2013). All three data sets indicate that Student Centered classes include specific strategies that serve to (1) engage students with one another; (2) engage instructors and students in a way that facilitates instructors assessing student learning and adjusting lessons accordingly; and (3) include questions, asked by both instructor and student, that determine the focus of a lesson.

Faculty who teach Student Centered classes have higher frequency and duration of engaging students with one another through in-class activities (e.g., small group discussions, think-pair-share activities, and exercises that have students work with data). In addition, faculty in Student Centered classes engage with their students to facilitate learning by asking and answering questions and monitoring student progress. A hallmark of these classes is that instructors assess student progress in the classroom and adapt or adjust the lesson to accommodate that progress. Pedagogical strategies that incorporate questions between students and faculty are also common in Student Centered classrooms. Faculty in Transitional and Teacher Centered classrooms use these teaching practices less frequently or not at all. The combined interaction of students and faculty measured by these data sets is consistent with learning theory that predicts that situating students in the context of a community enhances their ability to engage with material as they develop their own knowledge and intellectual agency (e.g., Wenger, 2000; Umbach and Wawrzynski, 2005; Freeman et al., 2014).

The convergence of quantitative and qualitative RTOP observation data plus instructor surveys also indicates that the combined suite of data is a powerful way to characterize teaching practices in geoscience classrooms. While the training to use the RTOP instrument is a multistage process that requires observers to carefully use a detailed rubric consistently, the information derived from observations provides a valuable tool for quantitative characterization of geoscience classes. As such, the approach described here would be useful for other applications, such as measuring the impact of curriculum design and implementation projects or of professional development programs.

From the foundation of this work, future research in professional development and student learning can now be based on the knowledge of what is happening in geoscience classrooms and what needs exist. These data can be used as a baseline to compare post-professional development instruction, and as a starting point for studies in student learning. Similarly, instructors can examine common teaching strategies employed in all three instructional categories, reflect on their own teaching, and determine whether they want to adapt such strategies in their own teaching practice.

APPENDIX. INSTRUCTOR SURVEY QUESTIONS

The instructor survey was taken online by 203 of the 204 instructors observed with the Reformed Teaching Observation Protocol (RTOP) instrument. Questions were designed to elicit information about typical teaching behaviors and the ways by which faculty make decisions regarding teaching methods.

Demographic Data

- How many years have you taught at the college or university level (do not include any experience as a graduate teaching assistant)?
- What is the highest degree level you have completed?
 - Masters
 - Ph.D.
 - Other (specify)
- Which of the following best describes your current position?
 - Full professor
 - Associate professor
 - Assistant professor
 - Instructor or lecturer
 - Adjunct or visiting professor
 - Other (specify)
- Name of course
- Type of course
 - Introductory course
 - Course for majors
 - Graduate level
 - Other (specify)
- Name of observer
- Length of class period observed (in minutes)

Describing Your Teaching

The following questions pertain to the entire course including lecture, labs, and discussion sections.

- In the "lecture portion" of your course, please estimate the percentage of class time spent on student activities, questions, and discussion.
- In the lecture portion of your class, please indicate how frequently you used the following teaching strategies in teaching your most recent course. Please use a scale from 1 to 5, where 1 is "never" and 5 is "nearly every class."
 - Traditional lecture
 - Lecture with demonstration
 - Lecture in which questions posed by instructor are answered by individual students
 - Lecture in which questions posed by instructor are answered simultaneously by the entire class
 - Small group discussion or think-pair-share

- Whole-class discussions
 - In-class exercises
- What changes have you made in the teaching methods in your course within the past two years (check all that apply)?
 - I have not made any changes
 - Spent less time lecturing
 - Employed more demonstrations during lectures
 - Increased questioning of students during lectures
 - Added group work or small group activities
 - Spent more time on class discussions or small group discussions
 - Changed assessment tools or strategies
 - Added assignments, e.g., more writing
 - Other (specify)
 - What do you rely on to determine if your teaching is working? (check all that apply)
 - Experience and gut instinct
 - Performance on exams, quizzes, assignments
 - Students show up for class and appear to enjoy class
 - Level of student engagement in class
 - Conversations with students
 - End-of-class or mid-term evaluations or surveys
 - Other (specify)
 - How do you learn about new teaching methods? (check all that apply)
 - Professional meetings or workshops
 - Publications
 - Discussions with other faculty members in my department
 - Discussions with colleagues in other institutions
 - Online resources
 - My own research
 - How has the use of online resources positively impacted your teaching within the past two years? (check all that apply)
 - Increased the variety of the methods that I use
 - Increased my skill with a particular teaching method
 - Increased by confidence as a teacher
 - Increased my ability to assess student learning
 - Influenced the topics that I address in my course
 - Increased my knowledge of a particular topic
 - Other (specify)
 - How often do you use the Cutting Edge website?
 - Never, I did not know there was such a website
 - Never, but I know of the website
 - Rarely
 - Monthly
 - Weekly or more often
 - Would you be willing to be contacted for a follow-up study related to how the use of On the Cutting Edge resources may have influenced your teaching?
 - Yes, I would be willing to be contacted for a follow-up study
 - Perhaps, you can contact me but I would need to know more about the follow-up study
 - No, I am not interested

TABLE A1. EXAMPLE QUOTATIONS FROM OBSERVER COMMENTS FOR EACH OF THE 41 ITEMS (ORGANIZED INTO 6 THEMES) IN THE QUALITATIVE ANALYSIS CODEBOOK

Item	Example quote from observer comments
Questions: How questions are asked and answered by both the instructor and students	
Instructor asks questions of students	"Lots of questions asked by the...teacher to students (10)."
No and/or few questions asked by instructor	"No questions for students at any point in lecture." "One shout-out question during whole class."
Short wait time, first shout-out taken, or instructor answers own questions	"There was a bit of wait time after questions, but usually the first shout out response was taken."
Good wait time after questions	"Prof showed clear wait time and if no students answered he clarified the question and waited until someone offered an answer."
Students answer questions	"Students answered questions alone and in pairs."
Students ask questions or volunteer ideas	"Several questions posed by students during lecture that didn't come from direct instructor prompts."
No or few questions from students	"No students asked any questions during lecture."
Questions from students not solicited	"No real pursuit of student questions."
Instructor pauses for or solicits student questions	"Pauses were built into the lecture to allow students to ask questions for clarification."
Shout-out questions asked	"Students were asked a few shout-out types of questions."
Assessment: How the instructor assesses students and makes use of assessment results	
Instructor assesses students (new or prior knowledge)	"The instructor used multiple elements of abstractions (equations, cross section, two types of plots) to develop and assess student understanding."
Instructor reminds students of previous topics/prior knowledge	"The class session began with a brief review of where the class had been previously and the overview was designed to remind students of the key pieces of previous topics."
No assessment of students	"Students are reminded of previous lessons, but instructor doesn't assess knowledge."
Lesson adjustments based on student work or prior knowledge	"The lesson adjusts significantly for student input but isn't entirely student-directed."
Instructor doesn't make use of student input; no shift in lesson	"The instructor respected student input, but it didn't shift the lesson."
Interaction: The frequency of student-student and student-faculty interaction	
Lots of student-faculty interaction	"This class had a high proportion of student-faculty interaction."
No or little student-instructor interaction	"Little to no ... student-professor interaction." "There was no dialogue between students or instructor except for point of clarification."
No student-student interaction and/or conversation	"There was no student-student interaction during this class."
Student-student interactions, group work	"Students were asked a number of questions and encouraged to discuss with neighbors prior to volunteering answers."
What students are asked to do: Types of activities or thinking students are asked to do	
No or little lecture, all or most of class is activity	"Instructor set up class period with one long activity for students to work in groups and then report out."
Students not asked to do anything; passive	"Students took notes, were not asked to participate."
Students read graphs, maps, measurements, photos, use data	"Students were asked to interpret graphs and a map."
Students make hypotheses or predictions	"Students made predictions, then compared answers with map in book."
Students reflect on learning	"Students were asked to reflect on what they learned when the learning goals for the entire course were reviewed."
Students answer open-ended questions	"Class time consisted of students working together on an open-ended problem."
Discussion (all scales)	"The class included small group discussion and a class discussion."
Student exploration	"Plan is for student directed investigation with nothing but student-student interaction; nothing but student exploration of data in map formats."
Problem-solving	"There was small group negotiating and problem solving."

(continued)

TABLE A1. EXAMPLE QUOTATIONS FROM OBSERVER COMMENTS FOR EACH OF THE 41 ITEMS (ORGANIZED INTO 6 THEMES) IN THE QUALITATIVE ANALYSIS CODEBOOK (continued)

Item	Example quote from observer comments
What instructor does: Types of behavior or lesson structures the instructor uses	
Lecture, professor talking, traditional teaching style	"Very traditional lecture."
Agenda and/or learning goals presented	"Instructor shows five learning objectives at start of class and highlights first appearance of learning objectives on subsequent PowerPoint slides."
Connections made with real world or other disciplines (perhaps through analogies)	"Instructor made a few everyday analogies (Ice skater spins faster if arms are pulled in)."
Real data and/or example presented	"The instructor used data from a local watershed and made a deliberate effort to tie the concepts to real world phenomena."
Conceptual framework not clear or not logical	"Students didn't understand conceptual framework, possibly because they felt overwhelmed by the assignment."
Mistakes made by instructor or student mistakes not corrected	"A few minor mistakes noted."
Alternate interpretations and/or ideas not explored or solicited	"Students were not engaged in considering, discussing, or even prompted to think about differing ideas/interpretations of data."
Good scaffolding of content or activities	"Exercise thoughtfully integrated various types of information in a well-scaffolded scheme."
Instructor acts as facilitator	"Instructor circulated around room during the activity and acted as a resource, but also to keep students on task and within the format."
Instructor circulates	"Instructor circulated through the room, asking/answering questions."
Engagement and overall mood: Types of student behavior with respect to engagement	
Students engaged	"The students are curious and engaged with the topic."
Some students not engaged; low-energy class	"There was very little energy in the room."
Students not taking notes; students doing other things	"Many students have laptops open and the computers are not on the class pages (open to email, Facebook, multiple web pages)."

ACKNOWLEDGMENTS

We are grateful to our team of 24 trained observers who spent hours in the training process and in completing the observations that formed the core data set for this project. We are also grateful to the 204 faculty who agreed to be observed for this project. The support of the Cutting Edge project leadership was instrumental in our ability to complete this work. The leadership team recognized this work as an important component of the larger Cutting Edge project, aimed at high-quality, ongoing professional development for geoscience faculty. This work is a direct outcome of their dedication to investment in a strong community of geoscience educators. Funding for this work comes from National Science Foundation Division of Undergraduate Education grant 1022844. We also appreciate thorough reviews from *Geosphere* editors, Sehoya Cotner, and anonymous reviewers who helped to greatly strengthen this manuscript.

REFERENCES CITED

Bransford, J.D., Brown, A.L., and Cocking, R.R., 2000, Learning: From speculation to science, in Bransford, J.D., et al., eds., *How People Learn: Brain, Mind, Experience, and School*: Washington, D.C., National Academy Press, p. 3–18.

Budd, D.A., van der Hoeven Kraft, K.J., McConnell, D.A., and Vislova, T., 2013, Characterizing teaching in introductory geology courses: Measuring classroom practices: *Journal of Geoscience Education*, v. 61, p. 461–475.

Carnegie Classification, 2015, The Carnegie Classification of Institutions of Higher Education, <http://carnegieclassifications.iu.edu/> (retrieved February 2015).

Chang, T.S., Lin, H.H., and Song, M.M., 2011, University faculty members' perceptions of their teaching efficacy: *Innovations in Education and Teaching International*, v. 48, no. 1, p. 49–60, doi:10.1080/14703297.2010.543770.

Chism, N.V.N., 2007, *Peer Review of Teaching: A Sourcebook* (second edition): Bolton, Massachusetts, Anker Publications, 228 p.

Desimone, L.M., Smith, T.M., and Frisvold, D.E., 2010, Survey measures of classroom instruction: Comparing student and teacher reports: *Educational Policy*, v. 24, p. 267–329, doi:10.1177/0895904808330173.

Dohaney, J., Brogt, E., and Kennedy, B., 2012, Successful curriculum development and evaluation of group work in an introductory mineralogy laboratory: *Journal of Geoscience Education*, v. 60, p. 21–33, doi:10.5408/10-212.1.

Ebert-May, D., Derting, T.L., Hodder, J., Momsen, J.L., Long, T.M., and Jardeleza, S.E., 2011, What we say is not what we do: Effective evaluation of faculty professional development programs: *Bioscience*, v. 61, p. 550–558, doi:10.1525/bio.2011.61.7.9.

Ebert-May, D., Derting, T.L., Henkel, T.P., Middlemis Maher, J., Momsen, J.L., Arnold, B., and Passmore, H.A., 2015, Breaking the cycle: Future faculty begin teaching with learner-centered strategies after professional development: *CBE Life Sciences Education*, v. 14, p. 1–12, doi:10.1187/cbe.14-12-0222.

Eddy, S.L., and Hogan, K.A., 2014, Getting under the hood: How and for whom does increasing course structure work?: *Cell Biology Education*, v. 13, p. 453–468, doi:10.1187/cbe.14-03-0050.

Eddy, S.L., Converse, M., and Wenderoth, M.P., 2015, PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes: *CBE Life Sciences Education*, v. 14, no. 2, p. 1–16, doi:10.1187/cbe.14-06-0095.

Epting, L.K., Zinn, T.E., Buskist, C., and Buskist, W., 2004, Student perspectives on the distinction between ideal and typical teachers: *Teaching of Psychology*, v. 31, p. 181–216, doi:10.1207/s15328023top3103_5.

Fairweather, J., 2009, Linking evidence and promising practices in science, technology, engineering, and mathematics (STEM) undergraduate education: A status report from the National Academies National Research Council Board of Science Education (BOSE): Washington, D.C., BOSE, http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_072637.pdf (accessed 9 June 2015).

Freeman, S., Eddy, S.L., McDonough, M., Smith, M., Okoroafo, N., Jordt, H., and Wenderoth, M.P., 2014, Active learning increases student performance in science, engineering, and mathematics: *National Academy of Sciences Proceedings*, v. 111, p. 8410–8415, doi:10.1073/pnas.1319030111.

Fung, L., and Chow, L.P.Y., 2002, Congruence of student teachers' pedagogical images and actual classroom practices: *Educational Research*, v. 44, p. 313–321, doi:10.1080/0013188022000031605.

- Gibbs, G., and Coffey, M., 2004, The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students: *Active Learning in Higher Education*, v. 5, p. 87–100, doi:10.1177/1469787404040463.
- Goldsmith, D.W., 2011, A case-based curriculum for introductory geology: *Journal of Geoscience Education*, v. 59, p. 119–125, doi:10.5408/1.3604824.
- Good, T., and Brophy, J., 1997, *Looking in Classrooms* (seventh edition): New York, Longman Press, 509 p.
- Greer, L., and Heaney, P.J., 2004, Real-time analysis of student comprehension: An assessment of electronic student response technology in an introductory Earth science course: *Journal of Geoscience Education*, v. 52, p. 345–351, doi:10.5408/1089-9995-52.4.345.
- Haak, D.C., HilleRisLambers, J., Pitre, E., and Freeman, S., 2011, Increased structure and active learning reduce the achievement gap in introductory biology: *Science*, v. 332, p. 1213–1216, doi:10.1126/science.1204820.
- Hake, R.R., 1998, Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses: *American Journal of Physics*, v. 66, p. 64–74, doi:10.1119/1.18809.
- Hallgren, K.A., 2012, Computing inter-rater reliability for observational data: An overview and tutorial: *Tutorials in Quantitative Methods for Psychology*, v. 8, p. 23–34, doi:10.20982/tqmp.08.1.p023.
- Henderson, C., and Dancy, M., 2007, Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics: *Physical Review Special Topics Physics Education Research*, v. 3, 020102, doi:10.1103/PhysRevSTPER.3.020102.
- Hora, M.T., 2015, Toward a descriptive science of teaching: How the TDOP illuminates the multidimensional nature of active learning in postsecondary classrooms: *Science Education*, v. 99, p. 783–818, doi:10.1002/sce.21175.
- Hora, M.T., and 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*, v. 22, p. 212–257, doi:10.1080/10508406.2012.729767.
- Kane, R., Sandretto, S., and 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*, v. 72, p. 177–228, doi:10.3102/00346543072002177.
- Knight, J.K., and Wood, W.B., 2005, Teaching more by lecturing less: *Cell Biology Education*, v. 4, p. 298–310, doi:10.1187/05-06-0082.
- Kober, N., 2015, *Reaching students: What research says about effective instruction in undergraduate science and engineering*: Board on Science Education, Division of Behavioral and Social Sciences and Education: Washington, D.C., National Academies Press, 256 p., doi:10.17226/18687.
- Kortz, K.M., Smay, J.J., and Murray, D.P., 2008, Increasing learning in introductory geoscience courses using lecture tutorials: *Journal of Geoscience Education*, v. 56, p. 280–290, doi:10.5408/1089-9995-56.3.280.
- Kuh, G., Kinzie, J., Schuh, J., and Witt, E., 2005, *Student success in college: Creating conditions that matter*: San Francisco, Jossey-Bass, 416 p.
- Lawson, A., et al., 2002, Evaluating college science and mathematics instruction: A reform effort that improves teaching skills: *Journal of College Science Teaching*, v. 31, p. 388–393.
- Lund, T.J., Pilarz, M., Velasco, J.B., Chakraverty, D., Rosploch, K., Undersander, M., and 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*, v. 14, no. 2, p. 1–12, doi:10.1187/cbe.14-10-0168.
- Macdonald, R.H., Manduca, C.A., Mogk, D.W., and Tewksbury, B.J., 2005, Teaching methods in undergraduate geoscience courses: Results of the 2004 On the Cutting Edge survey of U.S. faculty: *Journal of Geoscience Education*, v. 53, p. 237–252, doi:10.5408/1089-9995-53.3.237.
- Maclsaac, D., and Falconer, K., 2002, Reforming physics instruction via RTOP: *The Physics Teacher*, v. 40, p. 479–485, doi:10.1119/1.1526620.
- Manduca, C.A., Iverson, E.A., Czujko, R., Macdonald, R., Mogk, D.W., Tewksbury, B.J., McLaughlin, J., Sandord, C., Greenseed, L., and Luxenberg, M., 2011, Impact of the On the Cutting Edge professional development program on U.S. geoscience faculty: *American Geophysical Union Fall Meeting*, 2011, abs. ED42A-04.
- Markley, C.T., Miller, H., Kneeshaw, T., and Herbert, B.E., 2009, The relationship between instructors' conceptions of geoscience learning and classroom practice at a research university: *Journal of Geoscience Education*, v. 57, p. 264–274, doi:10.5408/1.3544276.
- Marlowe, B.A., and Page, M.L., 2005, *Creating and Sustaining the Constructivist Classroom*: Thousand Oaks, California, Corwin Press, 168 p.
- Mayer, D.P., 1999, Measuring instructional practice: Can policymakers trust survey data?: *Educational Evaluation and Policy Analysis*, v. 21, p. 29–45, doi:10.3102/01623737021001029.
- McConnell, D.A., Steer, D.N., and Owens, K.D., 2003, Assessment and active learning strategies for introductory geology courses: *Journal of Geoscience Education*, v. 51, p. 205–216, doi:10.5408/1089-9995-51.2.205.
- Michael, J., 2006, Where's the evidence that active learning works?: *Advances in Physiology Education*, v. 30, p. 159–167, doi:10.1152/advan.00053.2006.
- Multon, K.D., 2010, Interrater reliability, in Saldkin, N.J., ed., *Encyclopedia of Research Design*: Thousand Oaks, California, SAGE Publishing, p. 627–629, doi:10.4135/9781412961288.n194.
- Postareff, L., Lindblom-Ylance, S., and Nevgi, A., 2007, The effect of pedagogical training on teaching in higher education: *Teaching and Teacher Education*, v. 23, p. 557–571, doi:10.1016/j.tate.2006.11.013.
- Prince, M., 2004, Does active learning work? A review of the research: *Journal of Engineering Education*, v. 93, p. 223–231, doi:10.1002/j.2168-9830.2004.tb00809.x.
- Roehrig, G.H., and Kruse, R.A., 2005, The role of teachers' beliefs and knowledge in the adoption of a reform-based curriculum: *School Science and Mathematics*, v. 105, p. 412–422, doi:10.1111/j.1949-8594.2005.tb18061.x.
- Sawada, D., Piburn, M.D., Judson, E., Turley, J., Falconer, K., Benford, R., and Bloom, I., 2002, Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol: *School Science and Mathematics*, v. 102, no. 6, p. 245–253, doi:10.1111/j.1949-8594.2002.tb17883.x.
- Science Education Resource Center, 2014, *Classroom Observation Project: Understanding and Improving Our Teaching*: <http://serc.carleton.edu/52104> (accessed April 2015).
- Singer, S.R., Nielsen, N.R., and Schweingruber, H.A., eds., 2012, *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*: Washington, D.C., National Academies Press, 282 p., doi:10.17226/13362.
- Smith, M.K., Jones, F.H.M., Gilbert, S.L., and 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*, v. 12, no. 4, p. 618–627, doi:10.1187/cbe.13-08-0154.
- Smith, M.K., Vinson, E.L., Smith, J.A., Lewin, J.D., and Stetzer, M.R., 2014, A campus-wide study of STEM courses: New perspectives on teaching practices and perceptions: *CBE Life Sciences Education*, v. 13, no. 4, p. 624–635, doi:10.1187/cbe.14-06-0108.
- Stage, F.K., Muller, P.A., Kinzie, J., and Simmons, A., 1998, *Creating learning centered classrooms: What does learning theory have to say?*: ASHE-ERIC Higher Education Report, v. 26, 151 p.
- Stahl, R.J., 1994, Using think time and wait time skillfully in the classroom: *ERIC Digest*, no. ED370885, 4 p., <http://files.eric.ed.gov/fulltext/ED370885.pdf> (accessed June 2015).
- Tabachnick, B.G., and Fidell, L.S., 1996, *Using Multivariate Statistics* (third edition): New York, HarperCollins, 880 p.
- Umbach, P.D., and Wawrzynski, M.R., 2005, Faculty do matter: The role of college faculty in student learning and engagement: *Research in Higher Education*, v. 46, no. 2, p. 153–184, doi:10.1007/s11162-004-1598-1.
- Wainwright, C., Morrell, P.D., Flick, L., and Schepige, A., 2004, Observation of reform teaching in undergraduate level mathematics and science courses: *School Science and Mathematics*, v. 104, no. 7, p. 322–335, doi:10.1111/j.1949-8594.2004.tb18251.x.
- Walkington, C., and Marder, M., 2013, Classroom observation and value added models give complementary information about quality of mathematics teaching, in Kane, T.J., Kerr, K.A., and Pianta, R.C., eds., *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project*: San Francisco, Josey Bass, p. 234–277, doi:10.1002/9781119210856.ch8.
- Wenger, E., 2000, Communities of practice and social learning systems: *Organization*, v. 7, p. 225–246, doi:10.1177/135050840072002.
- Wieman, C., and Gilbert, S., 2014, The teaching practices inventory: A new tool for characterizing college and university teaching in mathematics and science: *CBE Life Sciences Education*, v. 13, no. 3, p. 552–569, doi:10.1187/cbe.14-02-0023.
- Wieman, C., Perkins, K., and Gilbert, S., 2010, Transforming science education at large research universities: A case study in progress: *Change: The Magazine of Higher Learning*, v. 42, no. 2, p. 6–14, doi:10.1080/00091380903563035.
- Yuretich, R.F., 2004, The effects of course redesign on an upper-level geochemistry course: *Journal of Geoscience Education*, v. 52, p. 277–283, doi:10.5408/1089-9995-52.3.277.
- Yuretich, R.F., Khan, S.A., Leckie, R.M., and Clement, J.J., 2001, Active-learning methods to improve student performance and scientific interest in a large introductory oceanography course: *Journal of Geoscience Education*, v. 49, p. 111–119, doi:10.5408/1089-9995-49.2.111.