Characterizing Teaching in Introductory Geology Courses: Measuring Classroom Practices

D. A. Budd,[1,a] K. J. van der Hoeven Kraft,[2] D. A. McConnell,[3] and T. Vislova[4]

ABSTRACT

Most research about reformed teaching practices in the college science classroom is based on instructor self-report. This research describes what is happening in some introductory geology courses at multiple institutions across the country using external observers. These observations are quantified using the Reformed Teaching Observation Protocol (RTOP). A scoring rubric created to support consistent application of the 25 items on the RTOP yields very high inter-rater agreement over multiple observations throughout a 3 y period. Using the adapted RTOP instrument, 66 separate observations of introductory physical geology classrooms at 11 different institutions (four associate's colleges, three baccalaureate colleges, a master's university, and three research universities) were collected, and those observations indicate three categories of instruction: (1) teacher-centered, traditional lecture-dominated classrooms (RTOP < 30) with little student talk and minimal student activity beyond listening and note taking; (2) transitional classrooms with some activities involving brief student discussions centered around right/wrong answers; and (3) student-centered classrooms (RTOP ≥ 50) with considerable time devoted to active learning and student communications to promote conceptual understanding. The progression from teacher-centered to transitional and then to student-centered categories is incremental across all subscales of the RTOP instrument except for propositional knowledge (character of the lesson's content and instructor's command of the material), which only increases between teacher-centered and transitional categories. This means there is no single path to an active learning, student-centered introductory geology classroom. Such learning environments are achieved with a holistic approach to all aspects of constructivist teaching as measured by RTOP. If the instructor incorporates small changes in multiple aspects of their teaching from disseminator of knowledge to supporter of student learning, then the transition to a student-centered classroom becomes an approachable process. Faculty can also use the RTOP and rubric to guide course planning, promote self-reflection of their teaching, and assist in the peer evaluation of other's teaching. © *2013 National Association of Geoscience Teachers*. [DOI: 10.5408/12-381.1]

Key words: RTOP, classroom observation, instruction, pedagogy, student-centered learning

INTRODUCTION

Science, technology, engineering, and math (STEM) instructors have access to many effective methods for improving learning in a range of introductory courses and disciplines (e.g., Ebert-May et al., 1997; Hake, 1998; Paulson, 1999; Crouch and Mazur, 2001; Wyckoff, 2001; Udovic et al., 2002; Crouch et al., 2004; Oliver-Hoyo et al., 2004; Knight and Wood, 2005; Singh, 2005; Beichner et al., 2007; Crowe et al., 2008; Kortz et al., 2008; Steer et al., 2009; Gray et al., 2010). These and other studies (Fairweather, 2009) consistently show greater student learning in classrooms that encourage students to analyze challenging questions, work collaboratively with small groups of peers, respond to instructor questions that assess learning, and focus on concepts over facts. These pedagogical strategies go by a range of names, but they generally fall under the banner of active learning. Courses utilizing active learning strategies are becoming more common in the geosciences and other STEM disciplines, but these changes are less commonly reflected in medium- to large-sized introductory geoscience classrooms (with more than 30 students), where fewer than 10% of instructors reported using active learning techniques (Macdonald et al., 2005).

Classroom observation protocols and self-assessment surveys provide systematic assessment of active learning classrooms. In various ways and to different degrees, these tools objectively assess the extent to which instruction is interactive, student-centric, and aligned with the American Association for the Advancement of Science (AAAS, 1990) definitions of constructivist teaching (i.e., the active engagement of the learner in the development of knowledge instead of a rote memorization forced upon the learner; Bransford et al., 2000). These tools provide a quantitative measure of classroom pedagogy that independent observers can apply across classrooms at different universities, which in turn allows investigation of a variety of research questions. For this study, we used an observation tool to assess two related questions regarding introductory physical geology classes. (1) To what extent are active learning teaching practices employed in introductory geology courses in American colleges and universities? (2) How do teaching practices differ between introductory geology classrooms that use and do not use active learning approaches?

We chose the Reformed Teaching Observation Protocol (RTOP; Piburn et al., 2000; Sawada et al., 2002) to help answer these questions. The RTOP instrument is aligned

Received 12 November 2012; revised 10 June 2013; accepted 8 July 2013; published online 22 November 2013.

[1]Department of Geological Sciences, 399 UCB, University of Colorado, Boulder, Colorado 80309, USA
[2]Physical Science, Mesa Community College at Red Mountain, 7110 East McKellips Road, Mesa, Arizona 85202, USA
[3]Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, North Carolina 27695, USA
[4]Department of Geology, University of St. Thomas, 2115 Summit Ave., St. Paul, Minnesota 55105, USA
[a]Author to whom correspondence should be addressed. Electronic mail: budd@colorado.edu.

TABLE I: Subscales of the RTOP instrument (from Sawada et al., 2002).

Subscale	Description
1. Lesson design and implementation	Assesses design and application of a lesson. Evaluates how the instructor organizes the lesson to honor students' preconceptions from other classes and everyday experiences. Are there opportunities for students to explore content before formal instruction? What is the intended role of the social construction of knowledge? Is student input used to focus and direct the lesson?
2. Propositional knowledge	Characterizes the lesson's content and the instructor's command of the material. Does the lesson highlight fundamental concepts? How clearly are concepts presented to illustrate the relationships among key components? Does the lesson incorporate ways for students to represent abstract concepts? Is content integrated with other disciplines and real-world applications?
3. Procedural knowledge	Assesses the skills, tools, and strategies an instructor employs to support student learning. Evaluates what the instructor asks students to do in the classroom. Much of this subscale relates to scientific ways of knowing and if students are engaged in these processes.
4. Communicative interactions (student–student interactions)	Evaluates the number, type, and quality of interactions among students. What is the extent of student–student communication and negotiation of understanding with peers? To what extent do students control their learning?
5. Student–teacher relationship	Appraises classroom culture and how the instructor promotes a culture of respect. Are students encouraged and comfortable asking questions? To what extent does the instructor help students with their activities?

with principles of constructivism and has well-established validity (Piburn et al., 2000; Sawada et al., 2002) and reliability (Sawada et al., 2002; Marshall et al., 2011; Amrein-Beardsley and Popp, 2012). It is one of the most widely used observation instruments in STEM college classrooms, having been employed by many researchers beyond its initial developers. (We found more than 40 studies of college-level instruction that employed the RTOP.)

Classroom observers using the RTOP instrument score each of 25 items on a five-point Likert scale (0 for "never occurred" to 4 for "very descriptive of the class"). Marshall et al. (2011) point out that one of the instrument's potential shortcomings is the interpretation of intermediate scores. That is, instructors or researchers studying RTOP data do not necessarily know the meaning of one instructor's score of 2 for an item relative to another instructor's score of 3 for the same item. To overcome this shortcoming, we developed a descriptive rubric that guides RTOP scoring and is applicable to all types of classroom environments and teaching styles. The rubric provides a framework for interpreting numerical values reported with RTOP scores and thus allows a robust characterization of classroom practices across institutions.

This paper presents the RTOP rubric developed for scoring the teaching and learning environment, reports the RTOP range and characteristics of 66 introductory physical geology classes, and discusses the two research questions posed in the introduction. The results reveal a broad spectrum in teaching strategies currently used in the classroom at college-level introductory physical geology courses. They suggest some current "norms" in terms of classroom characteristics and instructional pedagogy in these classes.

WHAT IS THE RTOP?

The RTOP (Piburn et al., 2000; Sawada et al., 2002) measures the degree to which reformed instructional practices are incorporated into lessons, thus shifting instruction from the traditional teacher-centered lecture-driven class to a student-centered, activity-based learning environment. The instrument builds upon inquiry and scientific reasoning tenets identified by the American Association for the Advancement of Science Project 2061: Science for All Americans (AAAS, 1990), the National Science Education Standards (NRC, 1996), and the Principles and Standards for School Mathematics (National Council of Teachers of Mathematics, 2000). The RTOP evaluates observable classroom processes, including the elements of lesson design and implementation, the content and processes of instruction, collaborations between students, and interactions between teachers and students. It consists of 25 items divided into five equal subscales (Table I).

The total score for the 25 items can range between 0 and 100, but most classes fall between scores of 20 and 80. Lower scores reflect traditional teacher-centered lecture classes, and higher scores represent student-centered, active learning environments (Sawada et al., 2002; Ebert-May et al., 2011). The RTOP has a high inter-rater reliability across classrooms and institutions (Sawada et al., 2002; Marshall et al., 2011; Amrein-Beardsley and Popp, 2012), and thus reliable comparisons can be made across classrooms within a single study. However, in the absence of a scoring guide, scores do not necessarily translate between studies. That is, a score of 40 derived by one team of observers in study A may not mean the same as a score 40 derived by another team of observers in study B.

The RTOP has multiple applications. It has been used to demonstrate significant student learning increases with greater implementation of student-centered active learning (Falconer et al., 2001; Lawson et al., 2002; Bowling et al., 2008; Budd et al., 2010). The RTOP also has been used as a peer evaluation tool (Amrein-Beardsley and Popp, 2012), for course design (Campbell et al., 2012), to assess the effectiveness of professional development programs (Adamson et al., 2003; Addy and Blanchard, 2010; Ebert-May et al., 2011), and as a standard to establish the concurrent validity of newer observation instruments (e.g., Erdogan et al., 2011; Marshall et al., 2011).

TABLE II: Values of Cronbach's alpha obtained for the RTOP and its subscales when scored with the rubric.

Subscale	Alpha
1. Lesson design and implementation	0.87
2. Propositional knowledge	0.36
3. Procedural knowledge	0.97
4. Communicative interactions	0.99
5. Student/teacher interactions	0.99
Entire RTOP	0.96

While the instrument's design ensures its applicability to different classrooms and objectives, it does have limitations. The RTOP does not assess any instructional or learning activity that occurs outside the classroom. Factors such as homework, associated laboratory classes, online resources, student attendance, and grading policy are not incorporated into the classroom observations. Lastly, no RTOP item focuses on the use of the content-specific learning goals.

METHODS

Development of the scoring rubric for the RTOP instrument is detailed in the online Supplemental Material (available at http://dx.doi.org/10.5408/12-381s1). Two observers (first and second authors) developed the rubric based on their own experiences teaching and observing introductory geology classes.[1] A score of 0 was accepted to mean that the item never occurred during an observation, and a score of 4 was described as a well-executed example of the respective RTOP item. Scores of 1, 2, and 3 were defined to capture the intermediate classroom processes or activities. Score descriptions were written to be independent of absolute number of students, subject matter of the lesson, equipment available, and physical arrangement of the classroom. The initial rubric was then tested and revised through a series of four classroom observations. Once finalized, 16 shared observations made between the fall of 2008 and spring of 2011 demonstrated excellent inter-rater agreement for total RTOP scores ($r = 0.940$; Fig. 1) and good inter-rater agreement for all 25 items ($r = 0.837$). In the few high scoring classrooms where differences in observers' total RTOP scores were >10% of each other, there were no items or subscales that consistently accounted for the discrepancy. For comparison, inter-rater agreements for total RTOP scores reported by other workers are 0.94 and 0.803 (Sawada et al., 2002), 0.83 (Roehrig and Kruse, 2005), 0.69 (Bowling et al., 2008), and >0.80 (Campbell et al., 2012). These comparisons suggest that the rubric provides greater scoring clarity, even for trained and calibrated observers.

As the RTOP construct itself was not changed, its validity (Piburn et al., 2000; Sawada et al., 2002) was assumed to be unaltered by the rubric. However, the

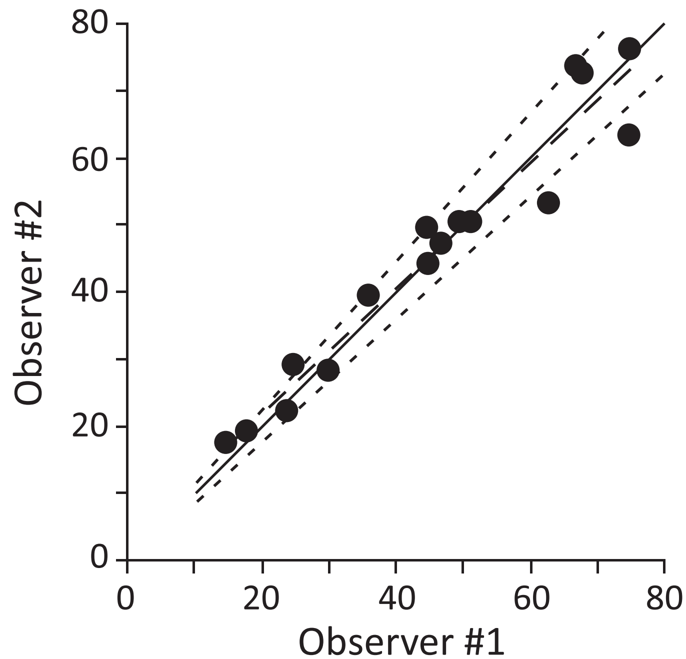

FIGURE 1: Correlation between RTOP scores in classrooms observed at the same time by both observers. Inter-rater agreement, as defined by the linear regression (long dashed line) correlation coefficient ($r = 0.94$), is excellent. Short dashed lines are ±10% of a one-to-one correspondence, which is the solid line.

addition of the rubric to the scoring requires reliability to be re-established. An instrument is considered reliable if it yields consistent results when used by different observers at different times (Roberson, 1999). We followed Sawada et al. (2002), Marshall et al. (2011), and Amrein-Beardsley and Popp (2012) and used Cronbach's alpha to assess the RTOP's reliability when scored with our rubric. Cronbach's alpha tests reliability by determining the internal consistency of items in a survey instrument (Cronbach, 1951; Santos, 1999). The higher the alpha, the more reliable is the instrument, with values >0.8 indicating good reliability and values >0.9 considered excellent reliability (George and Mallery, 2003). For the 16 sets of concurrent observer data collected for calibration, the standardized Cronbach alpha for the entire RTOP was 0.96 (Table II), which indicates that the rubric does not change the reliability of the RTOP instrument. With one exception, alphas for the subscales ranged from 0.86 to 0.99, which are similar to or greater than the alphas reported by Sawada et al. (2002), Marshall et al. (2011), and Amrein-Beardsley and Popp (2012). The exception is subscale 2, which has a standardized alpha of only 0.36 with all five items, and 0.60 if item 10 (connections to other disciplines and/or real-world phenomena were explored) is omitted. Item 10 was problematic because lower scores on it correlated to higher scores on items 6, 7, 8, and/or 9. Because of the unreliability of item 10 as a separate entity, it is not considered in the results, analysis, and discussion herein. Total RTOP score with item 10 included is used because the total RTOP's high alpha means item 10 does not affect the reliability of the overall instrument. Discussion of subscale 2 scores is considered without item

[1]The first and second authors developed the rubric and made all observations reported. One is a male with 24 y of teaching experience at a research university and a background in geoscience research, not education research. The other is a female with a background in geoscience education and 13 y teaching at a community college. The classrooms of both, plus the third coauthor, were observed; their scores are included in Table II, and they represent a range from transitional classrooms to student-centered classrooms.

10, since those scores also have a reasonable alpha once item 10 is removed.

All classroom observations were made as part of the GARNET project (Geosciences Affective Research Network; McConnell and van der Hoeven Kraft, 2011; Gilbert et al., 2012). Twenty-six instructors at 11 different institutions participated. To ensure that observations were characteristic of each instructor's teaching practices, all participants agreed to at least two classroom observations, and repeat observations were made in different semesters whenever possible. Nine of the participants were investigators on the GARNET project, and 13 additional instructors were recruited from the institutions of the GARNET investigators with no consideration other than a willingness to be observed. In order to increase the number of observations at associate and baccalaureate colleges, instructors known to the GARNET investigators were contacted, and four additional participants were recruited. Collectively, the participants represent four associate's colleges, three baccalaureate colleges, one master's university, and three research universities (classified as per Carnegie Foundation, 2010).

The participant pool is a sample of convenience chosen to test the RTOP rubric and provide a quantitative snapshot of teaching practices in introductory physical geology classrooms. Whether they are representative of the teaching practices of the national population of geoscience faculty is unknown because no characterization of classroom practices in that national pool exists. Ten of the 26 (38.5%) have been engaged in science education research or attended "On the Cutting Edge" workshops, which is higher than the approximately 25% of faculty from geoscience departments across the country reported to have participated in at least one "On the Cutting Edge" workshop (McLaughlin, 2009).

Sixty-six RTOP observations were made in introductory physical geology classes taught by the 26 instructors. Instructors included both new and highly experienced teachers and academic ranks range from part-time instructor to full professor. Observations were made between October 2008 and April 2012 by the same individuals who developed the rubric. Only lecture periods were observed; no associated recitations or laboratory classes were viewed. Nine of the 26 instructors were observed at different times by both observers. Each observation was arranged in advance with the instructor, but the instructors did not see the rubric or RTOP in advance. Observers sat in the midst of students and took observation notes during the class. If the physical arrangement of the room allowed, the observer moved amongst students during any prolonged activity and listened to student conversations. If movement was not feasible, the observer just listened to nearby students. After the class ended, the observer scored the RTOP using their notes and rubric.

Statistical analyses of the data were performed using SPSS, version 20. The Mann-Whitney *U*-test (also known as Wilcoxon rank-sum) was used to assess whether median RTOP scores of different demographic groupings were statistically different. One-way analysis of variance (ANOVA) was used to determine if the means of RTOP item scores in groupings of instructors were statistically appropriate. Due to the uneven numbers of observations within each grouping, homogeneity of variance could not be assumed for the ANOVA, so Welch's *F* statistic was used for analysis (Maxwell and Delaney, 2004). Because ANOVA only determines if there are significant differences between item scores, a follow-up analysis of Dunnett's *C* determined the statistical relationships between the groups. Effect size was also calculated to determine if the statistically reported differences were meaningful.

RTOP OBSERVATION RESULTS

Twenty-six different introductory physical geology instructors were observed (Table III). Instructors consisted of 10 females and 16 males, with a range of 1 to 29 y of teaching experience (median of 12 y). Twelve of the instructors teach at research universities, three at a master's university, seven at associate's colleges, and four at baccalaureate colleges. Class sizes clustered in ranges of 16–55, 72–90, and 121–168 students. For data analysis, the first of those clusters was considered small classes, and the other two clusters were grouped together as large classes.

RTOP scores for the 26 instructors ranged from 18 to 87 (Fig. 2), with a median of 42. For 23 (88%) of the instructors, the difference between their highest and lowest scores was 17 or less (median range of 7), indicating reasonable consistency from class to class in RTOP scores for those instructors. However, three instructors exhibited differences of 30 to 51 between their highest and lowest scores, indicating major differences in their learning environments from class to class. In all three cases, much of the range was generated in subscales 1 and 4 due to differences in lesson design and the presence/absence of student–student interactions. For example, one of those instructors presented a standard lecture during the first observation but devoted the entire second class to a multifaceted small-group activity that challenged students to make, analyze, and interpret their own set of observations.

The topics covered in the 66 observations spanned a wide variety of physical geology subjects, with earthquakes (11 observations, six instructors), water (nine observations, six instructors), deformation (eight observations, seven instructors), sedimentary rocks (eight observations, five instructors), and climate (five observations, three instructors) the most common content areas observed. Topics observed two to four times were minerals and igneous rocks, plate tectonics, shorelines, glaciers, energy, volcanoes, and relative age dating. Seventeen of the 26 instructors were observed teaching at least two different topics (Fig. 2). The Mann-Whitney *U*-test indicates that RTOP scores for those instructors are not significantly different from the scores of the other nine instructors ($\alpha = 0.05$, $p = 0.12$). This suggests that the topic of the observed lecture was not a significant factor in RTOP scoring.

We observed systematic differences in RTOP scores as a function of instructor gender, type of institution, and class size (Table IV; Fig. 3A–C). The Mann-Whitney *U*-test indicates statistically significant, higher average RTOP scores ($\alpha = 0.05$) for instructors of smaller classes ($\leq$55 students) compared to those with larger ($\geq$72 students) classes ($p = 0.005$), for female instructors compared to their male counterparts ($p = 0.008$), and for instructors at non-research universities compared to those at research universities ($p < 0.001$). RTOP scores at the master's university, baccalaureate colleges, and associate's colleges were similar to each other (Table IV), but the small sample sizes precluded statistical comparison. The lower RTOP scores at research universities

TABLE III: Characterization of participating instructors.

Instructor, gender	Years teaching[1]	Institution	Class Size[2]	Average RTOP[3]	RTOP Scores	Observer(s)
1, m	28	Research university A	149	18	15, 19, 21	1
2, m	2	Research university A	157	19	16, 19, 23	1
3, m	21	Research university B	77	25	23, 25, 26	1, 2
4, m	24	Research university B	74	25	21, 25, 27, 28	1, 2
5, m	9	Associate college B	72	26	25, 26	2
6, m	1	Baccalaureate college A	25	27	25, 30	2
7, f	3	Research university A	168	27	19, 30, 32	1, 2
8, m	10	Research university A	166	29	25, 32	1
9, m	27	Research university A	160	33	26, 34, 38	1
10, m	5	Baccalaureate college C	52	36	31, 41	1
11, f	2	Research university A	121	38	38, 38	1
12, f	20	Master's university A	80	41	35, 40, 40, 41, 51	1, 2
13, f	10	Associate college B	55	42	36, 40, 50	2
14, m	5	Master's university A	44	42	37, 47	1, 2
15, m	24	Research university A	167	43	42, 45	2
16, f	4	Research university A	164	45	40, 49	1
17, m	16	Associate college D	32	48	48, 48	2
18, f	13	Associate college B	34	50	35, 65	2
19, f	11	Associate college C	37	51	48, 54	1
20, m	29	Research university B	92	50	41, 54, 55	1, 2
21, m	10	Baccalaureate college B	24	56	30, 81	1
22, f	13	Associate college A	16	62	43, 68, 76	1, 2
23, m	24	Research university C	82	63	52, 67, 69	1, 2
24, f	6	Master's university A	90	64	61, 67	2
25, f	11	Associate college B	24	65	63, 67	1
26, m	20	Baccalaureate college A	32	87	85, 89	1, 2

[1]Years teaching at the time of the first observation.
[2]Size of first class observed if more than two observations.
[3]RTOP averages are presented to two significant figures.

and for male instructors, however, are more likely a reflection of class size. That is, all classes observed at research universities were large, whereas only 21% of classes at other institutions were large. Similarly, twice as many males as females taught large classes. Low RTOP scores in large classes thus are driving down average scores for males and research university instructors. Large class size is a known barrier to the implementation of a student-centered classroom (e.g., Henderson and Dancy, 2007), and high scores on observation instruments like the RTOP are difficult to achieve in such settings (Wainwright et al., 2004; Ebert-May et al., 2011). However, the challenges associated with large classes can be overcome, as evidenced by two instructors with large classes (>80 students) having average RTOP scores >60 (Table III). Another factor that could be promoting low RTOP scores in large classes at research universities is a bias to research over teaching, and thus lack of time and reward structure for faculty to change teaching methods (Henderson and Dancy, 2007).

There were no statistically significant differences ($\alpha = 0.05$, $p = 0.35$) in RTOP scores for instructors who had been teaching for more versus less than the median number of 12 y (Table IV; Fig. 3D). This result contrasts with the observations of Ebert-May et al. (2011), who found that the variable of years teaching was a negative predictor of RTOP score. However, six of the 12 most experienced instructors in our study are actively involved in science education reform as researchers or through professional development. It is thus possible that those activities offset years teaching in our data.

For all 66 observations, all five subcategories of the RTOP positively covary with total RTOP score (Fig. 4), an observation in keeping with the construct validity of the instrument (Piburn et al., 2000). Correlations are particularly strong for subscales 1, 3, 4, and 5, and those subscales also exhibit wide ranges in scores (0 to >15 out of 20). There is far less spread in scores for subscale 2 (Fig. 4). These trends indicate that average total RTOP scores are most influenced by lesson design, procedural knowledge (what students did), and the amount of student–student and student–teacher interaction.

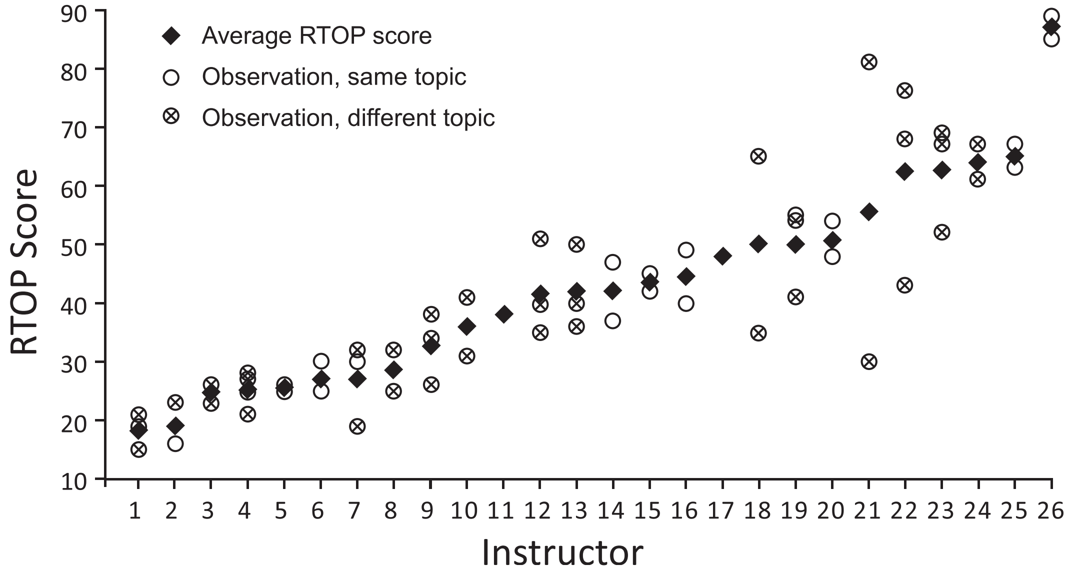

FIGURE 2: Distribution of average RTOP scores and scores of individual observations for all 26 instructors. Observations that covered the same topics (successive classes in one semester) are in open circles; all other observations were of classes that covered different topics (in the same or different semesters).

Instructors' RTOP scores can be grouped into three categories. Average RTOP ≤ 30 (eight instructors), average RTOP between 31 and 49 (nine instructors), and average RTOP ≥ 50 (nine instructors). We categorized these as: (1) teacher-centered, lecture-dominated classrooms where students are rarely talking; (2) transitional classrooms with some elements of active learning involving students talking, but with the sole purpose of seeking an answer; and (3) student-centered classrooms with more active learning that involves student talk to promote learning. Table V presents average scores for each RTOP item in each of these categories. ANOVA results show that the differences in the item averages are statistically significant, with a large effect size (Welch's $F = 81.95$, degrees of freedom [2, 37.2], $p < 0.001$, R^2 [effect size] = 0.69). Dunnett's *C* indicates that all three groupings are also statistically different from one another at $p < 0.05$.

Item averages (Table V) and differences in subscales scores (Fig. 5) as a function of RTOP category reveal four general patterns. First, the highest subscale score for teacher-centered classrooms is subscale 2, and transitional and student-centered instructors only exhibit slightly greater subscale 2 scores. This results, in part, from the fact that all instructors scored well (average >3.0) on item 8 (instructor had a solid grasp of the content inherent in the lesson) and item 9 (elements of abstraction were used when it was important to do so). This is not surprising, given that all observed instructors have graduate degrees in the geosciences and the diversity of imagery (outcrop and aerial photos, cross sections, maps, conceptual block diagrams, graphs, etc.) available to support student learning. Second, transitional and student-centered instructors record progressively higher scores in all of the other four subscales. The largest differences in subscale scores between instructor

TABLE IV: Population statistics for RTOP scores by demographic subgroups.

Demographic Category	Number of Instructors	Average RTOP score ($\pm 1\sigma$)	Median	Range
Females	10	47.1 ± 14.2	42	19–76
Males	16	37.8 ± 19.0	30.5	15–89
≤11 y teaching[1]	14	39.3 ± 16.1	37	16–81
≥12 y teaching[1]	12	43.4 ± 19.0	41	15–89
Research university	12	33.8 ± 14.6	30.5	15–69
All others	14	49.1 ± 12.8	47	24–89
Master's university	3	46.6 ± 11.1	41	35–67
Baccalaureate college	4	51.4 ± 28.3	36	24–89
Associate's college	7	49.4 ± 15.2	48	25–76
Class size ≥72 students	15	35.7 ± 14.6	33	15–69
Class size ≤55 students	11	51.5 ± 18.7	47.5	24–89

[1]Median years of teaching for all instructors is 12 y.

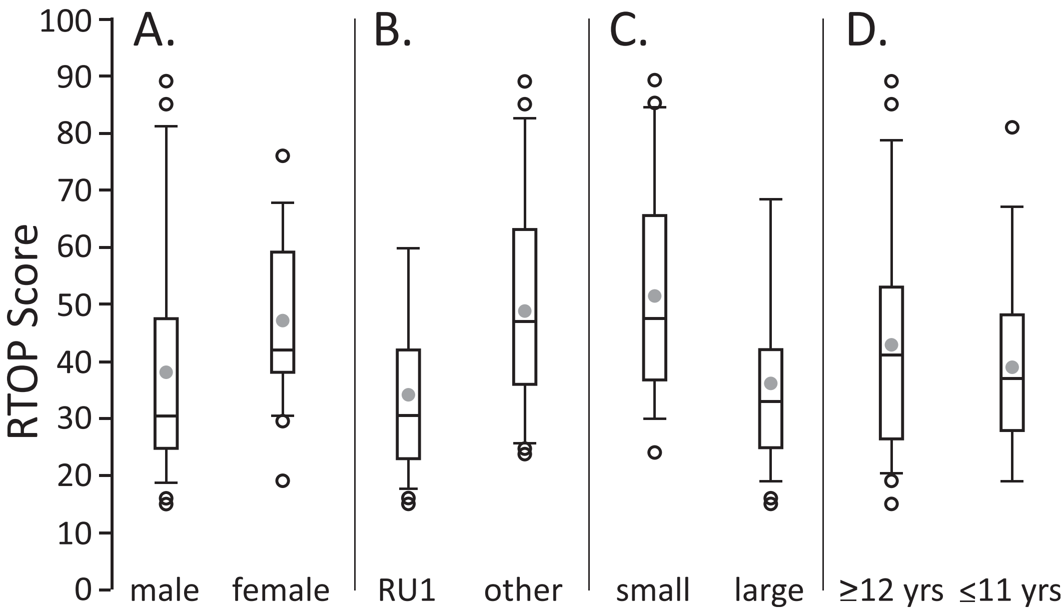

FIGURE 3: Whisker and box plots for RTOP scores as a function of instructor's (A) gender, (B) institution type, (C) class size, and (D) years of teaching experience (median is 12 y). Whiskers mark 10th and 90th percentiles; top and bottom of the boxes mark 75th and 25th percentiles. Solid line in middle of the box is 50th percentile; solid gray dot is the mean value; open circles are outliers. RU1 denotes research university.

categories (i.e., greater than the median difference of 4.0) occur in student–student interactions (subscale 4) between teacher-centered and transitional instructors and in subscales 1, 3, 4, and 5 between teacher-centered and student-centered instructors (arrows in Fig. 5). Third, 13 individual items account for most of the differences between traditional teacher-centered lecture and student-centered, active learning environments. Averages for those 13 items increase by ≥1.6 between teacher-centered and student-centered RTOP categories (Table V). Twelve of the 13 occur in subscales 1 (lesson design and implementation), 4 (student–student interactions), or 5 (student–teacher relations). Fourth, two items exhibit low average scores (1.3) in the student-centered category and are thus the greatest challenges to all introductory geology instructors. These are item 4 (lesson encouraged students to seek alternative modes of investigation or problem solving) and item 14 (students were reflective about their learning). Item 4 was scored >2 on only 9% of all 66 observations, which was the lowest percentage of all 25 items.

DISCUSSION

General Characteristics of Introductory Physical Geology Classrooms

The broad range in RTOP scores for introductory physical geology classes at a wide variety of institutions illustrates the value of the RTOP instrument for characterizing geoscience classrooms. Subscale scores and the rubric provide unique insight into the nature of physical geology instruction, characterize the constructivist steps that some instructors have successfully implemented, and define the most difficult instructional practices to implement. Vignettes based on observer notes from multiple classrooms illustrate the differences between teacher-centered, transitional, and student-centered classrooms.

Traditional Teacher-Centered, Lecture-Dominated Classrooms

Students slowly enter the classroom; the instructor is setting up a PowerPoint presentation. At 9 am, he dims the lights and starts to talk. Students quiet down. Some turn on their laptops and start taking notes; others are on Facebook, viewing email, or texting.

"Today we're going to continue talking about earthquakes. Last time I talked about faults; today I'll start talking about seismic waves." The instructor goes on to describe different types of seismic waves through a series of PowerPoint slides. After defining P-waves, S-waves, Love waves, and Rayleigh waves, he then asks for a student volunteer. A student raises his hand, and the instructor invites him to the front of the class. "Now don't let go of this spring or you'll regret it." Some student laughter, some students look up from their computer screens and smile. The instructor goes on to demonstrate the different seismic waves he just defined. The student volunteer sits back down. "You'll want to make sure you can distinguish between these different types of waves on the test next week." The instructor starts to describe how seismographs measure these seismic waves. A student raises her hand. The instructor calls on her and she asks, "Does this help us figure out how big an earthquake is?" The instructor replies, "That is exactly where I'm going" and then continues to describe how seismic waves are measured and how they behave differently. With a few minutes to go in the period, many students start packing up. The instructor responds, "I'm going to stop there. Next time I'll pick up with what these energy waves tell us about Earth's interior and intensity of earthquakes."

This vignette shows a teacher-centered classroom with an instructor who is well organized, has a thematic framework, and uses demonstrations to support student

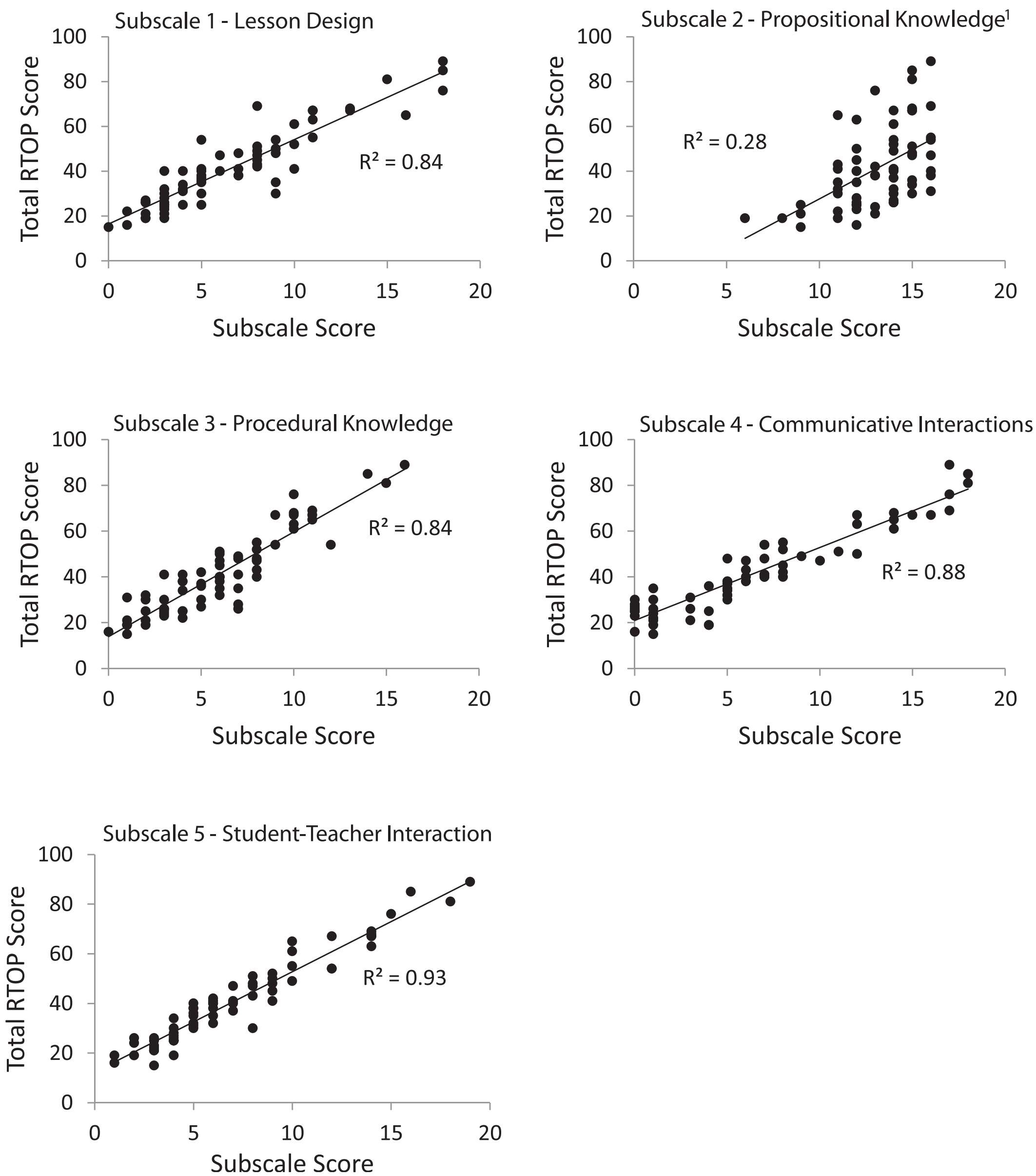

FIGURE 4: RTOP subscale scores versus total RTOP scores for all 66 observations. Subscale 2 has a maximum possible score of only 16 because item 10 is excluded due to its low reliability.

learning. However, the instructor does most, if not all, of the talking and thinking. The focus is on detail, covering material, and moving forward. The instructor possesses the knowledge and uses class time to convey his/her knowledge to the students. Students are inactive; there is no effort to determine if their minds are focused on the content. The instructor appears to assume that transmitting information equates to students learning content.

Teacher-centered classrooms score well on subscale 2 (propositional knowledge; Fig. 5) because instructors know the content and illustrate it using conceptual images, pictures, and demonstrations. Subscale 2 is in fact >50% of their total RTOP scores (Table V). However, even in this subscale, there may be shortcomings. The conceptual focus may not be clearly stated or emphasized, which may result in unclear connections between content and concepts. Terms and definitions (e.g., anticline, syncline, symmetrical and asymmetrical folds, plunging folds) are emphasized as much as fundamental concepts (e.g., compressive stress causes folding).

Subscale 1 scores are low (Fig. 5) because the lesson plan is designed to merely cover content. The instructor accommodates students' questions or comments but does not use student ideas to help guide the direction of the lesson. There is no plan for students to explore the content or concepts prior to the presentation. The instructor sets the stage by reminding students what was covered previously, which might include defining a basic concept for which the instructor assumes the students have an existing conceptualization (e.g., density, stress, convection). However, students are not asked to recall or engage their own prior knowledge.

Subscales 3 and 5 score low (Fig. 5) because students do not work with material, they are not asked to think deeply about the material, and student–teacher interactions are superficial. Questions to and from the instructor are the sole

TABLE V: Average RTOP scores for each item and subscale by instructional category.[1,2]

Item Subscale	Average, All Observations ($n = 66$)	Teacher-Centered Instructors' Average RTOP ≤ 30 ($n = 22$)	Transitional Instructors' Average RTOP 31–49 ($n = 23$)	Student-Centered Instructors' Average RTOP ≥ 50 ($n = 21$)
1. Lesson design				
1	1.9	**1.0**	**2.0**	**2.8**
2	1.9	**1.0**	**2.0**	**2.9**
3	0.8	**0.0**	**0.4**	**1.9**
4	0.6	0.1	0.3	1.3
5	1.5	**0.6**	**1.4**	**2.4**
Subscale Total	*6.7*	*2.7*	*6.1*	*11.3*
2. Propositional knowledge				
6	2.8	2.0	3.1	3.3
7	3.0	2.5	3.3	3.3
8	3.8	3.7	3.9	3.8
9	3.3	3.0	3.6	3.4
10	–	–	–	–
Subscale Total[3]	*13.9*	*11.2*	*13.9*	*13.8*
3. Procedural knowledge				
11	1.4	0.7	1.3	2.1
12	1.3	0.7	1.4	1.8
13	1.9	1.3	1.7	2.7
14	0.7	0.5	0.4	1.3
15	0.8	**0.1**	**0.4**	**2.0**
Subscale Total	*6.1*	*3.3*	*5.2*	*9.9*
4. Communicative interactions				
16	1.3	**0.3**	**1.3**	**2.2**
17	0.9	**0.2**	**0.7**	**1.9**
18	1.5	**0.4**	**1.7**	**2.5**
19	1.4	**0.4**	**1.6**	**2.2**
20	1.3	**0.4**	**1.4**	**2.2**
Subscale Total	*6.4*	*1.7*	*6.7*	*11.0*
5. Student–Teacher interactions				
21	1.6	1.0	1.5	2.4
22	0.7	0.2	0.3	1.6
23	2.0	**1.0**	**1.8**	**3.1**
24	0.8	**0.2**	**0.6**	**1.8**
25	2.1	**0.9**	**2.3**	**3.1**
Subscale Total	*7.2*	*3.3*	*6.5*	*12.0*

[1]Differences in all averages between the three categories are significant at $p < 0.05$.
[2]Bold font denotes the 13 items with an increase of 1.6 or more between low and high RTOP categories.
[3]With item 10 excluded due to low reliability, maximum possible for subscale 2 is 16, not 20.

vehicle for any student activity or student–teacher interaction. Instructors willingly answer students' questions, but they typically do not seek questions beyond the ineffective "any questions on that before I move on," with a wait time of mere seconds before continuing. This does not give students a chance to organize their comprehension, let alone frame a question. Questions are posed for individual students to answer (Why might that be? Does anyone know/remember? What do you think?), or as simple clicker questions for the whole class (e.g., term recall, restating content, identifying a geologic feature). For the former, the instructor takes the first raised hand, shouted answer, or may answer their own question, which eliminates the need for the majority of the students to actually consider the question. For clicker questions, students are typically quiet, either because they are not encouraged to discuss the

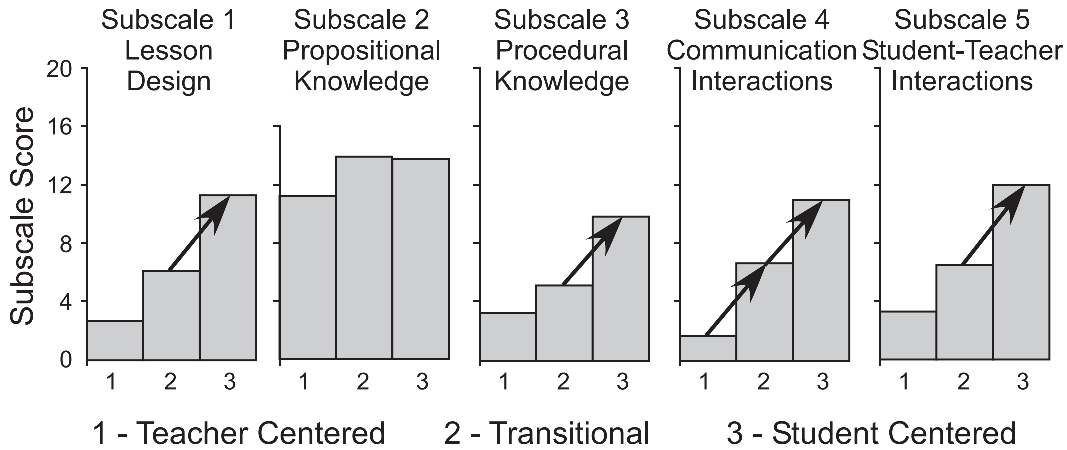

FIGURE 5: Bar charts of subscale scores as a function of instructor grouping. Arrows indicate greatest changes in subscale scores between instructor grouping (i.e., >median difference of 4.0). Subscale 2 has a maximum possible score of only 16 because item 10 is excluded due to its low reliability.

question with peers or because the question does not require discussion. The instructor is also quiet and does not move amongst the students to overhear their thinking or act as a resource to aid their thinking. The opportunity for students to interact with each other is limited to nonexistent, and thus the lowest scoring subscale is communicative interactions (Fig. 5).

First Steps to Active Learning—Transitional Classrooms

Students slowly enter the classroom; the instructor is setting up. She has two fault blocks, some string, a seismometer on a cart, and a PowerPoint presentation. At 9 am, she presents the first slide, "Outline for the day," which consists of learning goals. Students quiet down, some turn on laptops to start taking notes; others are on Facebook, viewing email, or texting.

"Today we'll be talking about earthquakes, but before we get started, I want to see what you remember from last time. Paul, do you remember what it's called when two blocks are moving past one-another? The instructor has two blocks and is demonstrating the movement for the student. Paul doesn't have a response, so she says, "Can anyone help out Paul? Feel free to use your notes." Another student shouts out, "a fault." "That's right! Why do we care about faults relative to earthquakes?" Carol raises her hand and answers, "Because that's where earthquakes come from?" "Good! So today we'll look at the energy they put out and how that can be measured." The instructor goes on to introduce the concept of seismic waves. At one point, she has students moving their arms to mimic compression and shear waves. When she describes how waves are recorded, she uses the string, a weight, and pen to illustrate a simple seismograph. She then projects a seismogram and explains how a modern seismograph works. Next is a clicker question, "What is the correct order of arrival time of the different seismic waves?" The instructor announces, "Talk to your neighbors." Many, but not all, students turn to the student sitting next to them and discuss the choices. "What do you think?" "I know it is not A." "I think it is D." "Why?" "I read last night that P waves were fastest." "Ok, I'll choose D too." After about 45 seconds, the instructor reveals the clicker responses. Nearly all students get the question correct. "Ok, why?" After a 10 second pause, "Come on, many of you answered this correctly." Four students raise their hands; she calls on one of them. "I recall reading P waves are faster." The instructor responds, "Good, reading the assigned text is helping us here. But why is it faster relative to the other wave types?" She then calls on one of the other raised hands, and that student responds with a suitable explanation. "Great! So know we know the sequence of wave types on all seismograms relates to how fast each wave type can move. Ok, so what does the seismogram tell us about magnitude of an earthquake?" After a short pause with no raised hands, she goes on to describe different scales and ways of measuring earthquakes. She tries to illustrate what magnitude means relative to an earthquake that occurred in the area a few years ago. She asks, "Did you feel that earthquake?" "What did you see happen?" Eight students raise their hand, and she calls on three who describe shaking lasting quite a while, items falling off shelves, and a pool sloshing back and forth for minutes. "Those observations are typical of what people reported, and they can be related to the scale of damage caused." The instructor writes additional information on the front board about effects reported in the local newspapers and relates those observations to points made in her lecture about seismic waves. At the end of class, she re-presents the outline, "OK, so this is what you should have learned today, be ready to pick it up from here next time."

This is an example of a transitional classroom in which instructors implement some elements of active learning. The instructor is still the dominant voice and thinker in the classroom, but student voices are now heard. Students talk to each other and the instructor, and there are efforts to engage student minds. Relative to teacher-centered classrooms, average RTOP scores improve for all subscales (Fig. 5), and they increase by a factor of two or more for subscales 1, 4, and 5 (Table V). These differences suggest that transitional instructors are distinguished from traditional instructors by design of their lesson, deliberate efforts to have students interact with each other, and the development of student–teacher interactions. The reformed teaching practices implemented are incremental and require only modest efforts on the instructor's part.

A major difference between teacher-centered and transitional classrooms is student involvement in the classroom. This involvement affects RTOP scoring across multiple subscales and constitutes the first steps to create a learning community. As illustrated in the vignette, the primary mechanism for student engagement is the instructor's questions to students. Asking questions beyond basic recall impacts the lesson design (subscale 1) and student–teacher interactions (subscale 5). Having students think and consider geologic information as part of the question impacts their procedural knowledge (subscale 3). Requiring students to talk to each other about the question affects communicative interactions (subscale 4).

Our observations indicate two to five questions to students per class period are typical for transitional classrooms. Some are simple recall questions; others are more challenging and ask students to interpret geologic information or make predictions. For example, rather than telling students what they should see in an image of a geologic feature, the instructor now asks them to describe what is shown, suggest how it got that way, or what it might mean relative to content discussed. Students will consider an image of a geologic feature (e.g., a fold, fault, rock, volcanic feature, hazard impact), or hypothetical scenarios described by text, block diagrams, or graphs. Transitional instructors allow students at least a small amount of time to contemplate problems and questions or allocate time for students to discuss the questions and learn from their neighbors. The instructor does not take the first shout out or first raised hand, an improvement over standard practice in traditional classes. However, the goal of questions or conversations is still to obtain a specific answer or line of reasoning. Student conversations and comments that are unrelated to the desired outcome are not allowed to change the direction of the class. The instructor is respectful of unsought answers but does not act on those ideas.

Concurrent with the implementation of some student activity and communication, lessons in transitional classrooms tend to be clearer and more logical with respect to concepts, and they involve fewer terms and definitions. For example, asking "what happened to these rocks?" (answer: they were folded by stresses squeezing them together) has greater conceptual focus than asking students to name the type of fold in an image. Transitional instructors also show more sensitivity to students' prior knowledge by asking, rather than telling, students what they learned in prior class periods (e.g., What did we do last time? What ideas did we have about . . .? What concepts were important when we discussed. . .?).

Second Steps—Achieving a Student-Centered Classroom

Students slowly enter the classroom. The instructor is setting up a presentation. At 9 am, he presents the first slide, "What do you already know about earthquakes?" The options are "nothing, a little, a lot, everything." Many students direct their conversations toward the question. The instructor gives students a minute to register their responses with clickers. Most select "a little." The instructor responds, "Ok, you know a little bit about earthquakes. Let's be more specific. Talk to your neighbors and make a list of what you know." While the students talk, the instructor moves around the room listening to conversations and commenting to different groups. After 4 min of discussion, the number of students turning to other activities is rising, so the instructor goes back to the front of the room. "I'm hearing a number of the same things in different groups. Who would like to share some of their thoughts on earthquakes?" Individuals voluntarily report out their group's ideas, and the instructor puts their responses into three lists on the board, each list reflecting a different theme that he expects to emerge. Some ideas are relevant, others less so, but all ideas are recorded. After 8 min, no group or individual has more to add, so the instructor points out how their ideas tie together. "What we're seeing is that you already know that earthquake intensity—its energy—is measured, they're caused by tectonic motion, and that they cause destruction (as he talks, the instructor places labels over each list of student thoughts). Now let's make connections between these ideas."

The instructor then gives a 16 min lecture on seismic waves and their relative travel times. He uses a rope, an animation, and a diagram to illustrate his points (he lumps all surface waves into one group). He has students "geogesture" with their arms in synch with the rope so that they personally visualize wave motions. The instructor then passes out copies of two seismographs. He states that these are recordings of seismic waves from the same earthquake but at two different locations. He explains the axes ("This direction is time since the earthquake, this axis is amount of ground motion—energy—detected at any one time") and projects the instructions on the screen. They are to work in small groups to (1) examine the seismographs and make a list of how they differ, (2) develop ideas as to why the two records might differ, and (3) note any confusion or surprises. The instructor moves about the room and helps some groups get started ("Compare these first two squiggles—were they recorded the same number of seconds after the earthquake?"), comments on other groups' thinking ("Yes, seems reasonable to me that all three types of waves should be detected at both locations"), and monitors progress. After ∼15 min, all groups are well into objective 2, but none have finished. The end of the period is approaching. He asks representatives from some of the groups to go to the board and simultaneously record their ideas on differences, choosing groups he knows to have different ideas. Once the information is on the board, he announces, "Finish this exercise before the next class. We will update the notes on the board and talk about your observations and ideas. Also think about how we might use the data on the seismographs to characterize the earthquake." Some students linger to argue which squiggly lines are the surface waves. The instructor photographs the board with his mobile device so he can reproduce the lists for the next class period.

In this student-centered classroom, the instructor has used several active learning activities and is using multiple strategies to engage students in an increasingly multifaceted learning community. Subscales 1, 3, 4, and 5 all show large scoring differences relative to transitional classrooms (Fig. 5 and Table V), indicating student-centric attributes in most classroom practices. The largest differences relate to lesson design and student–teacher interactions (subscales 1 and 5) as the instructor implements a role for themselves as the "guide on the side" (Sawada et al., 2002) rather than as the source of all learning. The instructor does not relinquish control of the classroom, and may still lecture, but students are explicitly charged with constructing their own understanding of the content. Instructors require students to explore before content is introduced, to activate their preexisting knowledge and conceptualizations, to work with and interpret data, and to communicate with each other and the instructor.

Activities that engage students (subscale 3) are one of the most distinctive aspects of the student-centered classroom. The activities are far more varied than in transitional classrooms. Small-group work involving questions or tasks at various cognitive levels is typical. Students are not just recalling and applying the content to a new situation as students in a transitional classroom might do (e.g., I lectured about fold types; can you identify this fold?). Instead, they are required to analyze situations or observations, make predictions, and/or compare and contrast ideas. Time spent on activities was observed to vary from student-centered class to class, but one-quarter of the class period was the general minimum. In most of our observations, the teacher framed questions and set up procedures to focus the students' work and thinking in a preferred direction. An instructor-led discussion of an activity typically followed its conclusion, during which groups or individuals shared their varied ideas and evidence, interpretations, and lines of reasoning.

The learning community that develops in student-centered classrooms is not dominated by student–teacher interactions, as in the transitional classrooms, but it includes an increasing role for interactions among students (subscale 4). Overwhelming evidence indicates that students learn best when they have opportunities to interact with one another and are not simply receivers of information (e.g., Bransford et al., 2000; Smith et al., 2009; Deslauriers et al., 2011). Interactions not only provoke students to form their own ideas and opinions, but also to consider using each other's ideas. We also observed students' conversations to go beyond just seeking an answer. To varying extents, they involved the negotiation of meaning and examination of problems in some depth. In smaller classes, we observed opportunities for every student to be heard and contribute.

Unlike the more rigid adherence to a lesson plan that is seen in teacher-centered and transitional classrooms, instructors in the student-centered classrooms exhibited flexibility (subscale 1). In the vignette presented here, the instructor was not bound to a rigid schedule. What was said and done by the students was far more important than attaining a predetermined stopping point. Flexibility also means that instructors listen to students and then act on what they hear. For example, when instructors assessed students' prior knowledge, they used what the students said to build the concepts and content for that day's class. In the higher scoring classrooms, instructors used students' ideas to direct the entire sequence of events in the class period.

If time is spent having students do things and talk to each other, the instructors' role as lecturer obviously diminishes, and less content is "covered" or transmitted. However, there are still lectures, with the instructor making a deliberate effort to focus on the most fundamental of concepts needed for the activities to succeed. The instructor's role changes in other ways too (subscale 5). While students worked with each other, whether for 1 or 45 min, the instructor aided students' thinking and interaction. Instructors showed great patience, suppressing any desire to tell students what they know and providing the time necessary to ensure the goals of the activity or conversation were achieved. The latter was even done in large classrooms, even though the instructor could not monitor all students equally.

The Greatest Challenges

Even in the student-centered, active learning classrooms, there are some RTOP items that show low average scores (Table V) and were rarely scored above a 2. These items are interpreted to represent the greatest challenges in introductory physical geology classrooms. They are the tasks that classrooms with total RTOP scores in the 60s and 70s are unable to accomplish.

Foremost among these challenges is item 4 (the lesson encouraged students to seek and value alternative modes of investigation or problem solving). This item relates to developing ways of thinking. The lesson plan in the most student-centered classes only called for the instructor (score 1; Supplemental Material) or students (score 2) to ask open-ended questions about investigative methods. Students did not engage in alternative modes of investigations (score 3), or discuss those alternatives (score 4). The only high scores on this item ($n = 3$) occurred when the entire class period was devoted to a single student activity. In those cases, the time necessary for students to decide how to proceed in the activity was not an issue. In all other cases, instructors prescribed how the activity was to be done, probably in part to ensure efficient use of time as the activity was to consume just part of the period. This suggests that the only way to improve scores on item 4 without using an entire period is to make an activity for which the sole purpose is to define how an investigation might proceed. This could be done via small group discussion and might require 5–10 min depending on the geologic phenomena the investigation is to explore.

Item 14 (students were reflective about their learning) was equally challenging to implement. In most observations, teachers did not create opportunities for students to be reflective. In the few cases where higher scores were given, students were provided time to reflect on what they had learned, but without much follow through on how those reflections connected to learning. Wainwright et al. (2004) and Flick et al. (2009) also found the use of metacognition strategies to be rare in science classrooms. We speculate that the rarity of reflection in introductory geology courses might occur for any of four reasons. First, the pressure to cover content drives instructors to forgo reflection. Second, instructors are so far along on the "expert" scale of knowing that they may have forgotten the value of reflecting about introductory material. Third, students are prompted to do reflection outside the classroom as part of a homework assignment, which is not captured by the RTOP. Fourth, instructors may have abandoned efforts to implement reflective activities if some students respond negatively to such activities. The value of promoting student reflection and metacognition in general, especially for those who underperform and lack insight into their shortcoming, has been well established (e.g., Schraw et al., 2006; Ehrlinger et al., 2008; National Research Council, 2012). Improving reflection will occur in any introductory classroom only if instructors explicitly ask students to reflect on their learning and provide class time to do so. For example, students may be asked to record their initial ideas about a topic before discussing the content (which also engages their prior knowledge). At the end of the topic, students could be asked to revisit their initial ideas and determine how they have changed, what has changed, and why they think change occurred (e.g., Kraft, 2012). Small group discussions can capture the key themes.

Applications of the RTOP

In addition to characterizing classrooms for research purposes, an instrument like the RTOP can be used to guide course planning (Campbell et al., 2012), promote self-reflection of teaching (MacIsaac and Falconer, 2002; Sawada et al., 2002; Wainwright et al., 2004; Addy and Blanchard, 2010; Amrein-Beardsley and Popp, 2012; Morrell and Schepige, 2012), assist in the peer evaluation of teaching (Amrein-Beardsley and Popp, 2012), and evaluate the impact of professional development and training (Adamson et al., 2003; Addy and Blanchard, 2010; Ebert-May et al., 2011). These applications need not be mutually exclusive. For example, self-reflection can lead to planning and implementing instructional changes followed by evaluation and re-reflection. The addition of the scoring rubric enhances the value of the RTOP in each of these applications, because the rubric, as opposed to a Likert scale, provides meaning to the spectrum of both micro (individual items) and macro (subscale) components of an instructor's classroom. The rubric is written in simple and straightforward terms, and thus it should be readily applicable and amendable by any user to these additional purposes. However, caution must be exercised, because RTOP scores are not valid unless those who are scoring have been appropriately trained. We thus emphasize using the instrument as a guideline in the examples outlined.

Course Planning

The RTOP scoring rubric can be used to guide course revision. Whether those reforms focus on a single subscale or draw from aspects of all subscales, the rubric provides both concrete examples of strategies that instructors can implement and a vision for what those strategies might look like in the classroom. For example, an instructor chooses to focus on increasing the interactions and communication between students (subscale 4). With the rubric as a guide, s/he decides to allocate 20% to 25% of the class period to student conversations that require students to work together, first in pairs and later in a larger group interaction. Given the time constraints, the instructor will not plan to use open-ended questions. Rather s/he designs tasks that require students to share and consider each other's ideas as they debate the meaning of some data, graph, or imagery. Students will not be charged to just seek answers, but will be instructed to apply concepts and decipher relationships. These plans, when scored using the rubric, yield a score of 9 for subscale 4, which is half way between the averages for a transitional and student-centered classroom (Table V). If the intent is to be more student-centered, then the rubric guides the instructor in ways to adjust their plan towards even more student engagement (i.e., a higher subscale score). A similar approach can be taken for any subscale or a specific item on the RTOP.

Self-Reflection

Reflection on one's teaching involves thinking about the processes of teaching and the reasoning therein (Kuit and Gill, 2001), and it can lead to the development of strategies that improve the learning environment (Boud and Walker, 1998). If the instructor's goal is active learning, then the RTOP provides a vehicle to guide the self-reflection process (MacIsaac and Falconer, 2002; Sawada et al., 2002). Scoring one's own lessons with the rubric provides an honest appraisal of what is actually happening in the classroom (Morrell and Schepige, 2012). Thinking about particular items or subscales (Is that something I do? How could I do it better? What would a more student-centered approach look like?) is facilitated by the structure of the RTOP and rubric. Amrein-Beardsley and Popp (2012) reported that faculty who used the RTOP found it most valuable in evaluating their own student–teacher interactions and communicative interactions.

Peer Evaluation

Peer evaluations of teaching in many institutions are conducted by departmental colleagues with limited background in pedagogy and no training in classroom observation. Evaluations thus might focus on just the mechanics of traditional teaching—assessing whether the instructor exhibited expert knowledge of the subject matter, gave a clear and well-organized presentation with appropriate supporting imagery, conveyed enthusiasm for the subject, and interacted with students (e.g., Yon et al., 2005). Without a framework for evaluation, assessment of even these mechanics may be vague. For example, does answering a single student's question demonstrate effective interaction? In contrast, use of the RTOP and rubric as a guide for their classroom observation requires the evaluator to consider more than the instructor's performance and propositional knowledge. It encourages the observer to examine multiple aspects of the teaching process and also provides context for evaluating the potential spectrum of classroom practices (Amrein-Beardsley and Popp, 2012). The final evaluation follows whatever protocol a department or college may have, but it is enriched by the RTOP-guided and rubric-calibrated observations. Equally importantly, the RTOP-guided observation can help peers with little experience in professional development to make constructive suggestions as to how colleagues might modify their teaching.

The RTOP could also be used as an assessment tool of professional development programs that focus on increasing student interactions or other reformed practices (Adamson et al., 2003; Addy and Blanchard, 2010; Ebert-May et al., 2011). Prior to any developmental training, the RTOP can be used to determine where an instructor's classroom practices lie on the teacher-centered to student-centered spectrum. Subsequently, the RTOP can be applied to measure whether instructors are implementing the best practices emphasized in their training, which also indirectly assesses the fidelity of the training program. Over time, repeated observations provide a longitudinal assessment of whether an instructor's classroom practices are evolving through multiple training experiences.

CONCLUSIONS

National surveys have indicated that an increasing number of geoscience faculty are self-reporting changes in their teaching practices that involve more active learning and student engagement (McLaughlin, 2009). Our observations of a small fraction of geoscience instructors demonstrate that some geoscience faculty are indeed stepping away from the lectern and talking to their students, encouraging students to talk and work with each other, and engaging students in classroom activities. Equally encouraging is the implementation of these constructivist strategies for student learning across a spectrum of class sizes, institution types, and years of teaching experience.

The presence of student-centered, active learning environments in introductory geology classrooms reflects the national trend of the instructor's role evolving from “talking head” to “learning coach” (National Research Council, 2012). To undertake a personal teaching evolution, geoscience faculty can use the RTOP and the scoring rubric presented in this paper to assess the current status of their classroom practices relative to teacher-centered lecture-dominated classrooms and student-centered learning environments. The results presented herein, combined with the RTOP and rubric, reveal insights and pathways any physical geology instructor can then follow in order to migrate their teaching to a more active learning classroom. No single type of intervention will foster the complete transformation. Rather, our results suggest change is occurring in classrooms through incremental steps related to lesson design, implementing a variety of learning activities, and fostering communication between everyone in the classroom. Using the RTOP rubric as a self-assessment tool will help faculty predict the cumulative effects of their plans with respect to the goal of a more active learning environment.

The RTOP and scoring rubric have other potential applications. As more faculty in a department consider reforming their teaching practices, tools like the RTOP can provide a common language for colleagues to use when discussing and evaluating the structure and delivery of courses (Wainwright et al., 2004). The RTOP and rubric have value as a research tool that can rigorously explore the link between classroom teaching practices and student learning (e.g., Falconer et al., 2001; Lawson et al., 2002; Adamson et al., 2003; Bowling et al., 2008), and the evolution of students' affect as a function of teaching style (e.g., McConnell and van der Hoeven Kraft, 2011).

Acknowledgments

This research was supported by the National Science Foundation Division of Undergraduate Education grants 0817260 and 1023097 to D. A. Budd, 0914404 to D. A. McConnell, and 1022980 to K. J. van der Hoeven Kraft. Dexter Perkins, Karl Wirth, and Ellen Iverson provided very useful comments on early drafts of the manuscript. We thank the instructors who opened the door to their classrooms and enabled us to conduct this research. We also thank Editor for Research Karen McNeal and two anonymous reviewers for their comments

REFERENCES

Adamson, S.L., Banks, D., Burtch, M., Cox, F., Judson, E., Turley, J.B., Benford, R., and Lawson, A.E. 2003. Reformed undergraduate instruction and its subsequent impact on secondary school teaching practice and student achievement. *Journal of Research in Science Teaching*, 40:939–957.

Addy, T.M., and Blanchard, M.R. 2010. The problem with reform from the bottom up: Instructional practices and teacher beliefs of graduate teaching assistants following a reform-minded university teacher certificate programme. *International Journal of Science Education*, 32:1045–1071.

American Association for the Advancement of Science. 1990. Science for all Americans. New York, NY: Oxford University Press.

Amrein-Beardsley, A., and Popp, S.E.O. 2012. Peer observations among faculty in a college of education: Investigating the summative and formative uses of the Reformed Teaching Observation Protocol (RTOP). *Educational Assessment, Evaluation and Accountability* 24:5–24.

Beichner, R.J., Saul, J.M., Abbott, D.S., Morse, J.J., Deardorff, D.L., Allain, R.J., Bonham, S.W., Dancy, M.H., and Risley, J.S. 2007. The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. *In* Redish, E.F., and Cooney, P.J., eds., Research-based reform of university physics. College Park, MD: American Association of Physics Teachers. Available at http://www.per-central.org/document/ServeFile.cfm?ID=4517 (accessed 13 March 2012).

Boud, D., and Walker, D. 1998. Promoting reflection in professional courses: The challenge of context. *Studies in Higher Education*, 23:191–206.

Bowling, B., Huether, C.A., Wang, L., Myers, M.F., Markle, G., Dean, G.E., Acra, E.E., Wray, F.P., and Jacob, G.A. 2008. Genetic literacy of undergraduate non-science majors and the impact of introductory biology and genetics courses. *Bioscience*, 58:654–660.

Bransford, J.D., Brown, A.L., Cocking, R.R., Donovan, M.S., and Pellegrino, J.W. 2000. How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

Budd, D.A., van der Hoeven Kraft, K.J., Knight, C., Wirth, K.R., McConnell, D., Bykerk-Kauffman, A., Matheney, R.K., Perkins, D., and Stempien, J.A. 2010. How much non-traditional teaching does it take to dramatically improve learning gains in introductory physical geology? *Geological Society of America Abstracts with Programs*, v. 42, p. 584. Available at https://gsa.confex.com/gsa/2010AM/finalprogram/abstract_181354.htm (accessed 26 October 2012).

Campbell, T., Wolf, P.G., Der, J.P., Packenham, E., and Abd-Hamid, N. 2012. Scientific inquiry in the genetics laboratory: Biologists and university science teacher educators collaborating to increase engagements in science processes. *Journal of College Science Teaching*, 41:82–89.

Carnegie Foundation. 2010. Carnegie Classification of Institutions of Higher Education. Available at http://classifications.carnegiefoundation.org/descriptions/basic.php (accessed 23 May 2013).

Cronbach, L.J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16:297–334.

Crouch, C.H., Fagen, A.P., Callan, J.P., and Mazur, E. 2004. Classroom demonstrations: Learning tools or entertainment? *American Journal of Physics*, 72:835–838.

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

Crowe, A., Dirks, C., and Wenderoth, M.P. 2008. Biology in BLOOM: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE-Life Sciences Education*, 7:368–381.

Deslauriers, L., Schelew, E., and Wieman, C. 2011. Improved learning in a large-enrollment physics class. *Science*, 332:862–864.

Ebert-May, D., Brewer, C., and Allred, S. 1997. Innovation in large lectures—Teaching for active learning. *Bioscience*, 47:601–607.

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*, 61:550–558.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., and Kruger, J. 2008. Why the unskilled are unaware: Further exploration of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, 105:98–121.

Erdogan, I., Campbell, T., and Abd-Hamid, N.H. 2011. The student actions coding sheet (SACS): An instrument for illuminating the shifts toward student-centered science classrooms. *International Journal of Science Education*, 33:1313–1336.

Fairweather, J. 2009. Linking evidence and promising practices in science, technology, engineering, and mathematics (STEM) undergraduate education: A status report for the National Academies National Research Council Board of Science Education (BOSE). Washington, DC: BOSE. Available at http://www7.nationalacademies.org/bose/Fairweather_CommissionedPaper.pdf (accessed 13 March 2012).

Falconer, K., Wycoff, S., Joshua, M., and Sawada, D. 2001. Effect of reformed courses in physics and physical science on student conceptual understanding. 82nd Annual American Educational Research Association Meeting, Seattle, WA. Available at http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED452083&ERICExtSearch_SearchType_0=no&accno=ED452083 (accessed 19 July 2012).

Flick, L.B., Sadri, P., Morrell, P.D., Wainwright, C., and Schepige, A. 2009. A cross discipline study of reformed teaching by university science and mathematics faculty. *School Science and Mathematics Journal*, 109:197–211.

George, D., and Mallery, P. 2003. SPSS for Windows step by step: A simple guide and reference, 11.0 update, 4th ed. Boston, MA: Allyn & Bacon. p. 357.

Gilbert L.A., Stempien, J., McConnell, D.A., Budd, D.A., van der Hoeven Kraft, K.J., Bykerk-Kauffman, A., Jones, M.H., Knight, C.C., Matheney, R.K., Perkins, D., and Wirth, K.R. 2012. Not just "rocks for jocks": Who are introductory geology students and why are they here? *Journal of Geoscience Education*, 60:360–361.

Gray, K., Steer, D., McConnell, D.A., and Owens, K. 2010. Using a student manipulated model to enhance student learning in a large lecture class. *Journal of College Science Teaching*, 40:86–95.

Hake, R. 1998. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66:64–74.

Henderson, C., and Dancy, M.H. 2007. Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physics Education Research*, 3:020102.

Knight, J.K., and Wood, W.B. 2005. Teaching more by lecturing less. *Cell Biology Education*, 4:298–310.

Kortz, K.M., Smay, J.J., and Murray, D.P. 2008. Increasing learning in introductory geoscience courses using lecture tutorials. *Journal of Geoscience Education*, 56:280–290.

Kraft, K.J. 2012. Encouraging students to think about how they learn. *In the Trenches*, 2:6–8.

Kuit, J.A., and Gill, R. 2001. Experiences of reflective teaching. *Active Learning in Higher Education*, 2:128–142.

Lawson, A., Benford, R., Bloom, I., Carlson, M., Falconer, K., Hestenes, D., Judson, E., Piburn, M., Sawada, D., Turley, J., and Wyckoff, S. 2002. Evaluating college science and mathematics instruction. *Journal of College Science Teaching*, 31:388–393.

Macdonald, R.H., Manduca, C.A., Mogk, D.W., and Tewksbury, B. 2005. Teaching methods in undergraduate geoscience courses: Results of the 2004 On the Cutting Edge survey of U.S. faculty. *Journal of Geoscience Education*, 53:237–252.

MacIsaac, D., and Falconer, K. 2002. Reforming physics instruction via RTOP. *The Physics Teacher*, 40:16–22.

Marshall, J.C., Smart, J., Lotter, C., and Sirbu, C. 2011. Comparative analysis of two inquiry observational protocols: Striving to better understand the quality of teacher-facilitated inquiry-based instruction. *School Science and Mathematics*, 111:306–315.

Maxwell, S.E., and Delaney, H.D. 2004. Designing experiments and analyzing data: A model comparison perspective, 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates. p. 868.

McConnell, D.A., van der Hoeven Kraft, K. 2011. The affective domain and student learning in the geosciences. *Journal of Geoscience Education*, 59:106–110.

McLaughlin, J. 2009. Evaluation Report of On the Cutting Edge. Available at http://serc.carleton.edu/files/NAGTWorkshops/2009_cutting_edge_evaluation_1265409435.pdf (accessed 3 June 2013).

Morrell, P.D., and Schepige, A.C. 2012. Helping preservice science teachers analyze their practices as we study our own. *Self-Studies of Science Teacher Education Practices*, 12:157–173.

National Council of Teachers of Mathematics. 2000. Principles and standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics.

National Research Council. 1996. Transforming undergraduate education in science, mathematics, engineering, and technology. Washington, DC: National Academy Press.

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

Oliver-Hoyo, M.T., Allen, D., Hunt, W.F., Hutson, J., and Pitts, A. 2004. Effects of an active learning environment: Teaching innovations at a research I institution. *Journal of Chemical Education*, 81:441–448.

Paulson, D.R. 1999. Active learning and cooperative learning in the organic chemistry lecture class. *Journal of Chemical Education*, 76:1136–1140.

Piburn, M.D., Sawada, D., Falconer, K., Turley, J., Benford, R., and Boom, I. 2000. Reformed Teaching Observation Protocol (RTOP). Available at http://physicsed.buffalostate.edu/AZTEC/RTOP/RTOP_full/ (accessed 1 October 2008).

Roberson, T.J. 1999. Classroom observation: Issues regarding validity and reliability. Available at http://www.eric.ed.gov/PDFS/ED427070.pdf (accessed 17 May 2012).

Roehrig, G.H., and Kruse, R.H. 2005. The role of teacher's beliefs and knowledge in the adoption of reform-based curriculum. *School Science and Mathematics*, 105:412–422.

Santos, J.R.A. 1999. Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of Extension*, 37:2TOT3. Available at https://cuvpn.colorado.edu/joe/1999april/,DanaInfo=www.joe.org+tt3.php (accessed 24 May 2012).

Sawada, D., 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*, 102:245–252.

Schraw, G., Crippen, K.J., and Hartley, K. 2006. Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, 36:111–139.

Singh, C. 2005. Impact of peer interaction on conceptual test performance. *American Journal of Physics*, 73:446–451.

Smith, M.K., Wood, W.B., Adams, W.K., Wieman, C., Knight, J.K., Guild, N., and Su, T.T. 2009. Why peer discussion improves student performance on in-class concept questions. *Science* 323:122–124.

Steer, D., McConnell, D.A., Gray, K., Kortz, K., and Liang, X. 2009. Analysis of student responses to peer-instruction conceptual questions answered using and electronic response system: Trends by gender and ethnicity. *Science Educator*, 18:30–38.

Udovic, D., Morris, D., Dickman, A., Postlethwait, J., and Wetherwax, P. 2002. Workshop biology: Demonstrating the effectiveness of active learning in an introductory biology course. *Bioscience*, 52:272–281.

Wainwright, C., Morrell, P., Flick, L., and Schepige, A. 2004. Observation of reform teaching in undergraduate level mathematics and science courses. *School Science and Mathematics*, 104:322–335.

Wyckoff, S. 2001. Changing the culture of undergraduate science teaching. *Journal of College Science Teaching*, 30:306–312.

Yon, M., Burnap, C., and Kohut, G. 2005. Evidence of effective teaching. *College Teaching*, 50:104–110.