

RESEARCH

Open Access



# Transforming education with community-developed teaching materials: evidence from direct observations of STEM college classrooms

Rachel Teasdale<sup>1\*</sup> , Katherine Ryker<sup>2</sup> , Karen Viskupic<sup>3</sup> , C. Doug Czajka<sup>4</sup>  and Cathryn Manduca<sup>5</sup>

## Abstract

**Background:** The Classroom Observation Project employs direct observations of geoscience teaching across the USA using the Reformed Teaching Observation Protocol (RTOP) to quantify the use of reformed teaching practices. We report on 345 RTOP observations used to evaluate the extent of teaching reform when curricular materials developed as part of the InTeGrate Project (ITG) were used. The InTeGrate Project has published 40 modules of curricular materials that teach geoscience in the context of societal issues and support instructors through guided use of student-centered instructional practices. All ITG materials were developed by teams of instructors, follow a consistent structure, and were evaluated against a project rubric.

**Results:** RTOP scores for classes observed when ITG materials were used (ITG;  $n = 50$ ,  $M = 54.0$ ) are significantly higher than RTOP scores for classes observed when ITG materials were not used (non-ITG;  $n = 295$ ;  $M = 39.8$ ;  $p < .0001$ ). ITG observations all have RTOP scores in the student-centered ( $\geq 50$ ) or transitional (31–49) instructional categories, and none in the teacher-centered instructional category ( $\leq 30$ ), demonstrating that ITG materials support more student-centered teaching in class sessions where they are used. In 33 paired observations of the same instructor teaching with and without ITG materials, mean RTOP scores when teaching with ITG are greater than mean RTOP scores when teaching without ITG ( $M = 54$  and  $M = 47.1$ , respectively).

**Conclusions:** RTOP observations reveal that more student-centered instructional practices occur in class sessions in which ITG materials are used. There is a small range of RTOP scores when individual ITG activities are used by multiple instructors, suggesting that using ITG materials results in a consistent quality of instruction. The complete absence of teacher-centered instruction when using ITG materials means the materials are a useful resource for practicing reformed teaching methods. The model of the ITG Project in the creation and broad dissemination of ready-made curricula for use in large numbers of classrooms can be replicated to transform teaching and learning in other disciplinary communities.

**Keywords:** Evidence-based instructional practices, Student-centered teaching, Reformed teaching, Observation protocol, Curriculum development, Geoscience or Geoscience education, Discipline-based educational research

\* Correspondence: [rtasdale@csuchico.edu](mailto:rtasdale@csuchico.edu)

<sup>1</sup>Department of Geological & Environmental Sciences, California State University, Chico, 400 W 1st St, Chico, CA 95929, USA  
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Introduction

Reformed teaching integrates active learning, problem solving, and critical thinking (Sawada et al., 2002), which prioritizes student agency, supports engagement (Umbach & Wawrzynski, 2005), and results in successful student learning (Freeman et al., 2014; Hake, 1998; Jang, Kim, & Reeve, 2012; Kuh, 2008; Prince, 2004). Reformed instructional practices shift instruction from a traditional teacher-centered, lecture-intensive class to a student-centered, activity-based learning environment (e.g., Budd, van der Hoeven Kraft, McConnell, & Vislova, 2013). Numerous calls have encouraged faculty to implement reformed, evidence-based instructional practices (e.g., NRC, 2012; Olson & Riordan, 2012) that are positively correlated with student learning in geoscience and other STEM courses (e.g., Freeman et al., 2014; McConnell, Steer, & Owens, 2003; National Academies of Sciences, Engineering, and Medicine [NASME], 2018; NRC, 2012; Olson & Riordan, 2012; Stes, Min-Leliveld, Gijbels, & Van Petegem, 2010). However, incorporation of reformed teaching strategies by STEM instructors has not been ubiquitous (e.g., Benabentos et al., 2020; Bradforth et al., 2015; Henderson, Beach, & Finkelstein, 2011; Stains et al., 2018; Teasdale et al., 2017; Wieman, Perkins, & Gilbert, 2010), especially in STEM classrooms (Keith, 2019).

In support of faculty adoption of student-centered teaching and to increase geoscience literacy in undergraduate students, new curriculum materials were developed through the InTeGrate project (ITG; Interdisciplinary Teaching about Earth for a Sustainable Future), which was funded by the National Science Foundation Science, Technology, Engineering, and Mathematics Talent Expansion Program (NSF-STEP) Center in the geosciences (InTeGrate, 2019a). Materials were designed to engage students in learning about Earth through the context of societal grand challenges while working with data and scientific practices (Egger, Bruckner, Birnbaum, & Gilbert, 2019). The ITG materials broaden the spectrum of freely available online geoscience materials (e.g., those available on Teach the Earth, <https://serc.carleton.edu/teachearth/>) because they address interdisciplinary, societally relevant problems, which may help a broader audience of students learn the principles of geoscience (Gosselin, Manduca, Bralower, & Egger, 2019). ITG materials are also uniquely designed according to a common set of standards to ensure that they are aligned with all goals of the project, including the use of evidence-based instructional practices (Steer et al., 2019). The reformed or student-centered teaching practices emphasized in ITG materials address the social construction of knowledge among students and between students and instructors (Vygotsky, 1978). The need for

ITG materials arose from the knowledge that teaching about societal issues is of central importance to the geosciences (marine, Earth, and atmospheric sciences), and that individual instructors would benefit from instructional resources that also incorporate evidence-based teaching practices (e.g., Henderson & Dancy, 2007; Sunal et al., 2001).

Undergraduate teaching has not been widely transformed to reflect the best practices for STEM undergraduate instruction (e.g., reformed teaching, student-centered instruction) identified by Discipline-Based Education Research (DBER) and cognitive science (e.g., Henderson, Mestre, & Slakey, 2015; Stains et al., 2018; Teasdale, Viskupic, et al., 2017). Even if STEM instructors are aware and convinced of the importance of such practices, a variety of barriers such as lack of training, time, and incentives can limit the use of reformed teaching practices (American Association for the Advancement of Science [AAAS], 2011; Brownell & Tanner, 2012; Henderson et al., 2011; Henderson, Finkelstein, & Beach, 2010; Marbach-Ad, Egan, & Thompson, 2015). A possible strategy for increasing the use of reformed, student-centered teaching is to provide instructors with curricular materials they can use in their courses that already include reformed teaching strategies. In doing so, instructors could efficiently gain experience with reformed teaching practices, which itself might serve as a form of professional development (PD). The widespread adoption of ITG materials, which ubiquitously use reformed teaching methods, provides an opportunity to evaluate if instructional materials can increase the use of reformed teaching in undergraduate courses. ITG materials have been adopted, adapted, or influenced the design of more than 3000 courses enrolling more than 100,000 students across the USA (Gosselin, Manduca, et al., 2019; Kastens, 2019).

The ITG materials are freely accessible, ready-to-use community resources. Assets within ITG materials include guidelines to help instructors use the materials, and descriptions of modifications or adaptations made by instructors to better fit their course context or teaching needs. These elements of the ITG resources are designed to help mitigate barriers to instructor use. While no set of curriculum materials can possibly overcome every obstacle, ITG materials address many of the most highly cited barriers to reform (e.g., time, support). Thus, evaluation of the teaching methods employed by faculty using the ITG materials allows a robust test of the role of materials in supporting improvements in teaching practice. If successful, the materials development, design, and assessment processes could be adapted for other projects (Egger et al., 2019).

## Research questions

The goal of our research is to measure the impact of ITG materials on teaching practices by answering the following research questions:

1. What are the effects of teaching with ITG on the use of student-centered instructional strategies?
2. Can anyone, regardless of their level of participation with ITG, teach with ITG materials and have a student-centered class?
3. Do different instructors using the same ITG materials teach with similar levels of reform?

These research questions were addressed by analysis of classroom observation data collected using the Reformed Teaching Observation Protocol (RTOP; Sawada et al., 2002) in classes in which instructors did and did not use ITG materials.

## InTeGrate teaching materials and materials development rubric

The InTeGrate project engaged the higher education community of geoscientists and social scientists in related disciplines to create interdisciplinary instructional materials to teach geoscience in the context of societal issues (Gosselin, Egger, & Taber, J. J. (Eds.), 2019). The ITG project published curricular materials as complete courses, and also as modules that could be used within new or established courses. All modules are divided into units (individual lessons) that can be used independently or in series. Most modules were designed to take approximately 2 weeks of time during a standard semester-long course that meets approximately three hours per week. ITG modules were designed to be used in lecture class sessions (but can also be used in laboratory sections), and do not require supplies other than what is available (e.g., downloaded) from the modules.

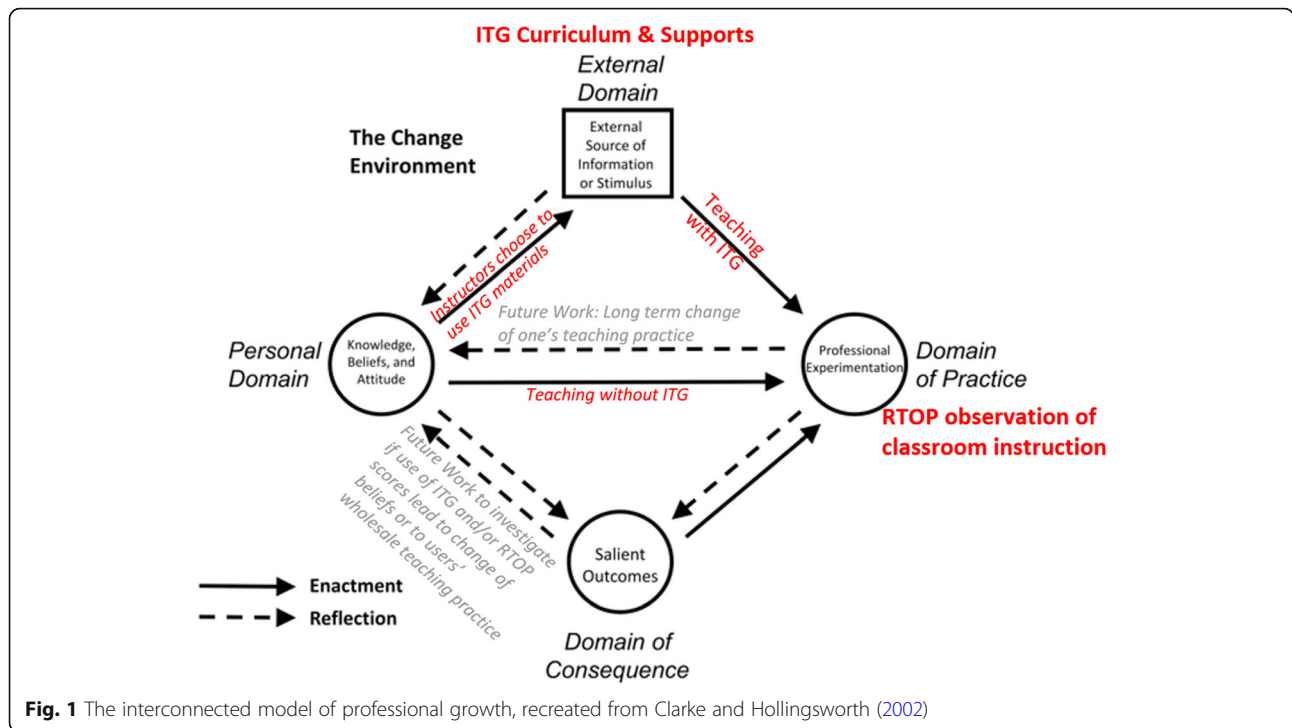
ITG materials were authored by interdisciplinary teams from different institution types (e.g., 2-year-, primarily undergraduate, research-, or comprehensive institutions). Each ITG unit includes learning goals, the context for use of the materials (e.g., information or content students should already know before starting the unit), and teaching materials. Teaching materials include step-by-step instructions for instructors, and most units include a pre-class assignment for students, an instructor's presentation (usually in PowerPoint), student handouts, data sets, teaching notes with suggestions for timing and logistics of activities, formative and summative assessments, and references and additional resources. Formative and summative assessments in ITG materials include answer keys that are only available to instructors. Additional support materials include instructor stories that describe how the module authors

used the materials in their courses. The instructor stories provide additional guidance for implementing ITG activities in different contexts, such as courses of different size, student demographics, and classroom setup.

All materials developed as part of the ITG project were evaluated against the InTeGrate materials development rubric, ensuring they were held to consistent standards in terms of alignment with the project's guiding principles and with research-based pedagogical best practices. The rubric has 28 elements distributed among six sections: overarching goals, learning objectives and outcomes, assessment and measurement, resources and materials, instructional strategies, and alignment. The ITG materials development rubric evaluated curricular materials as a whole; the rubric was applied at the module level and not the unit level. Final review of materials also included scientific, technical, and copyediting reviews for content and formatting accuracy. This review process ensured that all ITG materials met consistently high standards, including the incorporation of reformed teaching strategies. For a full description of the materials development rubric, see Steer et al. (2019).

## Theoretical framework

Changes to an individual's teaching practices are motivated and informed by varied factors, which may include changes in beliefs about the teaching and learning process (Addy & Blanchard, 2010; Anderson, 2002), conversations with colleagues about teaching (Riihimaki & Viskupic, 2019), professional development (Ebert-May et al., 2015; Manduca et al., 2017; Viskupic et al., 2019), or an increase in peer support and time to develop practices for a specific context (Austin, 2011; Fairweather, 2008; Wieman et al., 2010). Clarke and Hollingsworth (2002) described the environment in which change takes place as consisting of four interacting domains, which together make their interconnected model of professional growth (Fig. 1). The four domains are external (source of information or stimulus), practice (professional experimentation), consequence (salient outcomes), and personal (knowledge, beliefs and attitudes); these are interconnected through enactment and reflection. In the Clarke and Hollingsworth (2002) model, change in one domain that leads to change in another is a "change sequence." A change sequence that results in a lasting change, or professional growth, is a "growth network" (Clarke & Hollingsworth, 2002). We use Clarke and Hollingsworth's (2002) model to contextualize how ITG materials and their development can impact teaching practices (domain of practice) and how elements of the ITG materials development process could be extrapolated to the development of other curricular materials (Table 1, Fig. 1; see section on "Characteristics of



participants and instruments” for details on the different groups).

**Individuals involved in the change process**

A major goal of the ITG project was for materials developers to design curricular materials that could be used by any instructor to increase students’

geoscience literacy and engage students in the learning process (Egger et al., 2019). It was expected that the development of and access to the ITG materials would drive a change in teaching practices (professional experimentation), with the supports outlined in Table 1. Variations in how instructors interact with the RTOP-ITG project components described in

**Table 1** A summary of components of the RTOP-ITG project considered in this paper (white rows) that can support a change in teaching (Clarke & Hollingsworth, 2002) when instructors use ITG. Additional supports outside of the ITG project are also possible, but are not known for individual instructors as part of this study and are the topic of future work (grey rows)

Domain	RTOP-ITG project components	RTOP-ITG participants
External domain: external source of information or stimulus	ITG Materials, which include resources of: Guidance in the use of ITG materials (Teaching materials and instructions, Teaching notes and tips, instructor stories, instructional webinars) Participation in Research Team or Implementation Team, including PD and assessment of use of ITG materials (see section “Characteristics of Participants and Instruments”)	All instructors Mentored instructors
Domain of practice: professional experimentation	Process of creating materials, including workshops and webinars on materials development, collaboration during materials development, including evaluation with the ITG materials development rubric Measured by RTOP observations (Research Questions address how the enactment of the external domain of ITG impacts teaching)	Materials Developers (including ITG authors, module editors and assessment team) All observed instructors
Not examined in this paper		
Domain of consequence: salient outcomes	Student outcomes from assessments provided in ITG materials RTOP observation scores (by subscale, including information on how to interpret one’s scores)	All instructors All observed instructors
Personal domain: knowledge, beliefs and attitude	Personal reflections	All instructors

Table 1 may result in variations in the degree to which the domain of practice is affected.

An instructor might decide to use the ITG materials through many paths, which is part of the design of the ITG project (Egger et al., 2019; Kastens & Manduca, 2017; O'Connell, Bruckner, Manduca, & Gosselin, 2016). For example, an instructor may have experience teaching traditional oceanography and geology courses, but may want to more explicitly include the societal relevance of the disciplines, believing that their students will be more engaged (domain of consequence). If they watched an InTeGrate webinar (external domain), they may explore some of the materials on their own and identify Earth's Thermostat (Dunn, MacKay, & Resor, 2019) as a module to use. They may read and reflect on the instructor story (external domain) from one of the materials developers who used Earth's Thermostat in a similar course setting, which suggests the module would be similarly successful in their own course (personal domain). Rather than the traditional lectures on climate change this instructor had used in previous courses, they decided to try teaching with the ITG module.

The pathway for an observed instructor who uses ITG includes some knowledge or belief (personal domain) that motivates the use of ITG (external domain), which results in teaching practices that are measured with the RTOP (professional experimentation). An instructor who is observed when they do not use ITG bypasses the external domain (e.g., they were not aware of the materials or believe they were not appropriate to use in this class) and their teaching practices are also measured with the RTOP (professional experimentation). A change from one domain to another represents a possible change sequence for that class session (measured in this project to address RQ1-3), while a lasting change to one's teaching practice would represent a growth network (Clarke & Hollingsworth, 2002), which is not measured here. Our research questions explore the impact of ITG curriculum as a new external source of information (external domain) on instructors' teaching

during observed class sessions (as measured by RTOP; domain of practice).

## Methods

### Design and setting

To characterize the impact of using ITG materials on teaching practices, teaching was observed using the RTOP for a broad array of US college-level instructors when they were either using ITG materials (ITG) and/or when they were not using ITG materials (non-ITG). We compare the RTOP scores for observations of all instructors using ITG and all instructors not using ITG, and also analyze differences in observations of single instructors teaching both with and without ITG materials (RQ1, Table 1). We investigated the level of training needed to successfully use ITG materials by comparing instructors' professional development related to the ITG project with the instructional category observed when using the ITG materials (RQ2). In some cases, observations were made of different instructors using the same ITG unit within a module. This allowed us to examine the internal consistency of teaching practices (domain of practice) that are built into the design of ITG modules (RQ3). Table 2 provides a summary of research questions, instructor types and analyses.

### Characteristics of participants and instruments

#### Participants

There are several groups of people involved in the ITG project. Materials developers are teams of three or more people who wrote, piloted, and revised an ITG module together. Materials Developers worked closely with ITG module editors and assessment team members. A member of the assessment team consulted with each team of materials developers and was responsible for reviewing the developed curriculum against the ITG materials development rubric. Materials developers also wrote about their experiences with the ITG materials they developed in instructor stories; this provided a structured opportunity to reflect on the domain of practice and what it means for the salient outcomes of that instructor. All

**Table 2** Organization of this project's research questions and participation by 287 observed instructors

Research Questions	RQ1: What are the effects of teaching with ITG on the use of Student-Centered instructional strategies?	RQ2: Can anyone, regardless of their level of participation with ITG, teach with ITG materials and have a student-centered class?	RQ3: Do different instructors using the same ITG materials teach with similar levels of reform?
Participant types (see text for definitions)	a. All observations ( $n = 345$ ) b. Paired instructors (observed when teaching with and without ITG), $n = 33$	All paired instructors (materials developers, mentored users, unmentored users)	All ITG users (materials developers, mentored users, unmentored users)
Analysis	a. Welch's unequal variance $t$ test comparing scores of observations with ITG vs. scores of observations without ITG b. Two-tailed $t$ test	Two-way mixed ANOVA compares RTOP scores across three participation categories	Comparison of RTOP scores for ITG units taught by more than one observed instructor

See detailed demographics in Fig. 2 and types of participation in Fig. 3.



instructors using ITG materials could then refer to the instructor stories as sources of information.

Some instructors other than materials developers were mentored as part of the ITG implementation or research teams, and participated in various PD activities in the use of ITG materials. Implementation teams used the ITG materials at their own institutions (or groups of institutions), documenting how materials were adapted in diverse contexts (Orr, C.H. & McDaris, J.R., 2019) and participated in peer-guided PD specific to their individual projects. Research team members attended a 3-day workshop, and then used 18 class sessions of ITG materials in their courses (Czajka & McConnell, 2019). Unmentored instructors are users of ITG materials who were not part of an ITG project that provided PD (e.g. not members of an Implementation or Research team).

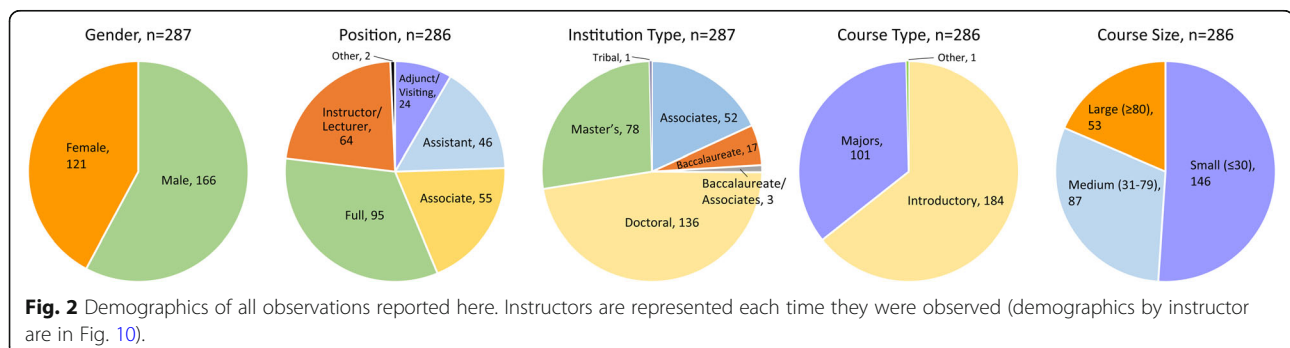
Observed instructors represent a convenience sample in that they agreed to be observed primarily by responding to an individual invitation by an RTOP observer. Additionally, 37 observations are of instructors who participated in ITG-related projects and were recruited through targeted emails. Efforts were made to include instructors across a range of demographic factors including geographic location, institution type, class size, and course level. The ITG materials were designed for interdisciplinary use and to encourage the incorporation of geoscience content into non-geoscience courses, and five ITG observations were made of non-geoscience faculty in non-geoscience courses (philosophy, history, religious studies, biology, and nursing). The 345 observations completed for this project include 287 unique instructors, and intentionally represent a broad spectrum of US college and university faculty (Fig. 2).

**Instructors' ITG participation categories** To investigate if the degree of reformed teaching in ITG classes was influenced by the level of instructor interactions with the ITG materials (i.e., differences in the external domain; RQ2), we assigned an ITG participation category to each instructor who was observed teaching with ITG (Iverson & Wetzstein, 2020): (1) materials developers; (2) mentored instructors; and (3) unmentored

instructors (Table 3 and Fig. 3). ITG assessment team members and ITG module editors had extensive experience working with ITG materials and a deep understanding of the ITG project goals and guiding principles, so they were included as materials developers in this study. Mentored instructors were part of an ITG Implementation team or the research team. Unmentored instructors were not part of a project focused on the use or development of ITG materials. Because they were not part of PD associated with ITG, unmentored instructors are hypothetically most representative of future ITG users who will access and use ITG materials.

### Instruments

**Classroom observations** The RTOP was developed specifically to measure the use of reformed teaching practices in K-12 classrooms (Piburn & Sawada, 2001; Sawada et al., 2002). However, the RTOP has been used in undergraduate classrooms to measure levels of reform in geoscience (Budd et al., 2013; Lund et al., 2015; Teasdale, Viskupic, et al., 2017) and other STEM courses (Lund et al., 2015). The RTOP has also been used to describe changes to instructors' teaching practice associated with professional development (PD) programs in STEM (e.g., Ebert-May et al., 2015; Lawson, Benford, Bloom, & Carlson, 2002; MacIsaac & Falconer, 2002). Similar research in the geosciences used a subset of data presented here to compare PD participation with RTOP scores for non-ITG observations (Viskupic et al., 2019). Total RTOP scores can range from 0 to 100, with higher scores resulting from more reformed, student-centered teaching practices. A standardized rubric was added to the RTOP and adopted to facilitate consistent scoring (Budd et al., 2013; Teasdale, Viskupic, et al., 2017). Total RTOP scores were used to assign each classroom observation to an instructional category according to the classification of Budd et al. (2013): teacher-centered (RTOP score  $\leq 30$ ), transitional (RTOP score 31–49), or student-centered (RTOP score  $\geq 50$ ). A team of trained observers used the RTOP and rubric to make observations of classroom teaching practices for this project.



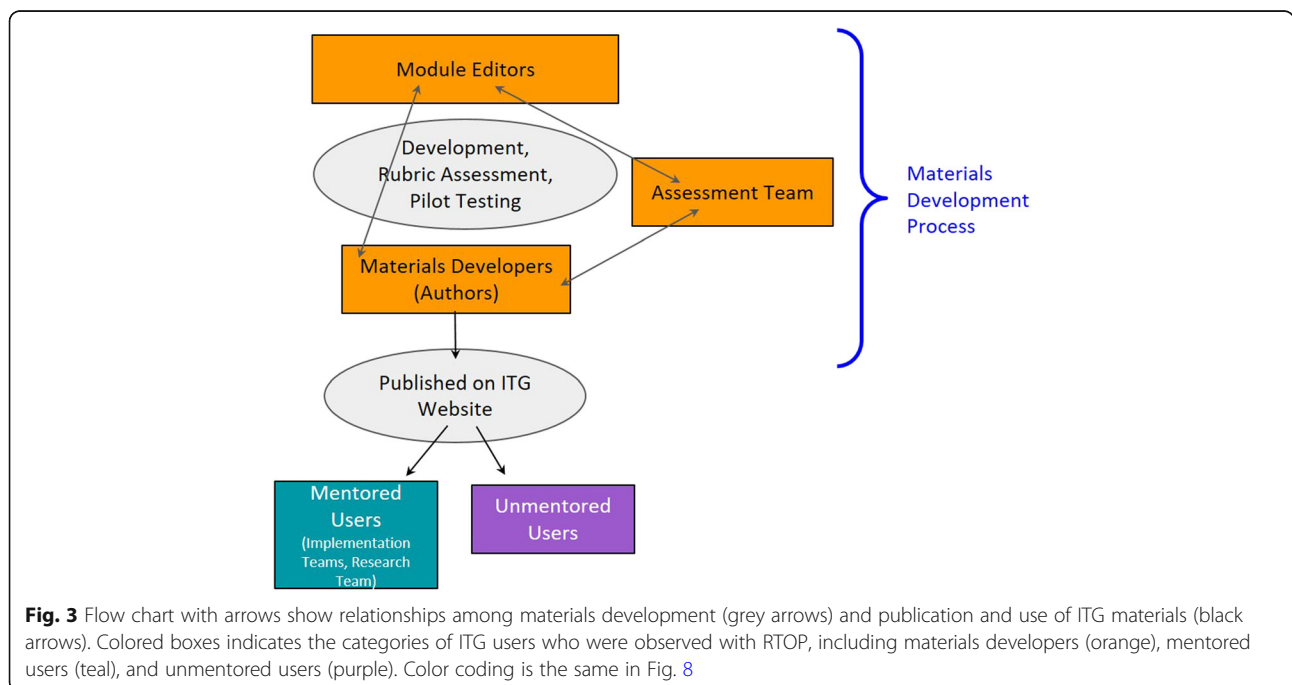
**Table 3** Observations of instructors by their ITG participation category

Participant types	Materials developers	Mentored instructors	Unmentored instructors
Description	ITG module authors—generally in teams of 3 faculty from different institution types (to ensure diverse teaching perspectives)	ITG users who participated in ITG-related projects such as implementation teams, research teams. Mentoring in the use of ITG and student-centered teaching practices occurred through organized professional development related to ITG projects	ITG users who were not authors or part of a project focused on the use or development of ITG-materials
Total observations $n = 345$	27	37	281
Number of ITG observations ( $n = 50$ )	13	21	16
Number of non-ITG observations ( $n = 295$ )	14	16	265
Number of paired observations $n = 33$ (both ITG and not-ITG)	10	12	11

The 25 RTOP rubric items are divided into five subscales: lesson design (design and application of the lesson), content—propositional knowledge (content and its organization in the lesson), content—procedural knowledge (what students do during the lesson), classroom culture—communicative interactions (types of interactions among students), and classroom culture—student-teacher relationships (types of interactions between the instructor and students). The subscales contribute to the total RTOP score (e.g., Sawada et al., 2002). Methods to assess instructional practices other than direct observations include self-report surveys (e.g., Wieman & Gilbert, 2014), interviews (e.g., Markley, Miller, Kneeshaw, & Herbert, 2009) and combinations of methods (e.g., Ebert-May et al., 2015; Teasdale, Viskupic,

et al., 2017). Direct observations are considered to provide robust and objective measures of classroom activities when a trained observer quantifies an observation using a rubric to systematically record teaching practices (e.g., RTOP; COPUS used by Smith, Jones, Gilbert, & Wieman, 2013; PORTAAL used by Eddy, Converse, & Wenderoth, 2015), or when observations are used to provide more descriptive data (e.g., TDOP of Hora, 2015).

The RTOP observation team was trained and calibrated using procedures described by Teasdale, Viskupic, et al. (2017) and Viskupic et al. (2019). Observers watched, scored, and discussed video recorded classes to become familiar with the RTOP rubric, and were required to pass successive stages of the training by



scoring videos to within one standard deviation of the accepted score (the mean score from calibrated observers). Annual calibration of all observers ensured that scoring standards remained constant during the 10-year period of observations that contributed data to this project. Observers using the RTOP have both high inter-rater reliability (e.g., Amrein-Beardsley & Popp, 2012; Marshall, Smart, Lotter, & Sirbu, 2011; Teasdale, Viskupic, et al., 2017) and high intra-rater reliability (Viskupic et al., 2019). Calibration of observers for this project resulted in a Cronbach's alpha of 0.996 (with  $\alpha$  of video 1 = 0.81 ( $n = 22$ ) and  $\alpha$  of video 2 = 0.84 ( $n = 24$ )), which exceeds the threshold for inter-rater reliability ( $\alpha > 0.7$ ; Multon, 2010; as reported in Teasdale, Viskupic, et al., 2017). Reproducibility of observations by individual observers (intra-rater reliability) was calculated for three observers on our team using the method of Bland and Altman (2003). All repeated observations fell within the 95% upper and lower limits of agreement with a standard deviation of 3.48 points (Viskupic et al., 2019).

Annual calibration of all RTOP observers ensured that scoring of observations made during the study were consistent. A Pearson's correlation indicates there is no significant change in RTOP non-ITG scores over time in our data set ( $r(292) = -0.021$ ,  $p = 0.724$ ). This suggests that it is appropriate to include all observations in our data set, as there has not been a widespread change in teaching practices as measured by the RTOP in non-ITG classrooms since data collection started.

The RTOP measures research-based best practices that contribute to student-centered teaching during an individual class session (Teasdale, Viskupic, et al., 2017; Viskupic et al., 2019). Previous work established that RTOP observations are consistent with instructor self-report of teaching practices (Ebert-May et al., 2015; Teasdale, Viskupic, et al., 2017), and that reformed teaching, as quantified by high RTOP score, correlates with student conceptual gains (e.g., Bowling et al., 2008; Lawson et al., 2002; MacIsaac & Falconer, 2002).

The 345 classroom observations discussed here were completed between April 2009 and May 2019. As part of each observation, instructors were asked to complete an instructor survey, provide contextual information to the observer about the class being observed (pre-observation interview), and allow the observer to sit in the classroom and apply the RTOP rubric without interacting with students or the instructor during the class period. Instructors received a post-observation report, indicating their score and some contextualized information regarding average scores for the project up to that point.

The 345 observations were made of 287 unique instructors. Most instructors (242) were observed only once using either ITG or non-ITG material. A smaller

sample of 46 instructors were observed multiple times (103 observations total) to (a) compare practices of the same instructor teaching with and without ITG materials, or (b) for other comparisons unrelated to the research questions addressed here (Czajka & McConnell, 2019; Macdonald et al., 2019).

Of the 46 instructors observed when teaching with ITG, four were observed multiple times using ITG, resulting in 50 total ITG observations. Thirty-three instructors were observed at least once each using ITG and non-ITG materials; observations of these instructors are referred to here as "paired." Among the 33 instructors for whom we have paired observations, seven were also observed twice when teaching without ITG and two were observed twice when teaching with ITG. Classroom observations are not intended to reflect an instructor's entire teaching practice, but are used as representations of an instructor's typical class session, which observers confirmed with instructors in the pre-observation interviews.

In cases of multiple observations during which instructors did not use ITG materials, only the earliest observation is used for comparisons because this observation likely represents the smallest influence of the ITG project on an instructor's teaching practices. Because the order of observations does not influence RTOP score (see Results section), selection of the first observation also eliminates potential bias that might occur by choosing an instructor's highest or lowest score.

In cases of multiple observations of the same instructor using ITG materials, only the earliest ITG observation is used for comparisons because this observation represents the instructor teaching with the least amount of experience using ITG materials. Whether or not the instructor was involved in professional development related to the use of ITG, the first observed use of ITG is likely more similar to other less experienced users' instruction with ITG.

**Instructor survey** Each observed faculty member completed an electronic survey to collect demographic information such as gender, position type (rank), course enrollment, and course level (e.g. introductory or majors). The instructor survey also asked questions to measure participants' typical teaching practices including use of different instructional activities (e.g., lecture, discussion, questioning strategies) and proportion of class time spent using such activities (Teasdale, Viskupic, et al., 2017).

#### Statistical analyses

Descriptive statistics are reported for the entire population and for the population of instructors who used ITG.



Inferential statistics appropriate for each research question were completed using IBM Statistical Package for the Social Sciences (SPSS) version 25. Each analysis is described in detail in the “Results” section and summarized in Table 2.

#### Statement on ethics approval and consent

Research protocols were reviewed and approved by the Institutional Review Board. All observed instructors in this study signed a consent form before their classroom observation. Student feedback was not included in this study and observers did not interact with students. In some cases, observed instructors chose to tell their students that an observer was present; in other cases, observers were not announced.

#### Limitations

Classroom observations using the RTOP were made during single class periods of in-person instruction, and did not include observation of any laboratory, field, or supplemental instructional periods. We do not suggest that a single observation captures an instructor’s entire teaching practice, but observations of instructors when not using ITG were scheduled to ensure that the class period observed would be “typical” of the instructor’s teaching practice, as described by the instructors themselves. Ebert-May et al. (2015) and Teasdale, Viskupic, et al. (2017) report agreement between results of classroom observations using the RTOP and instructors’ self-reported teaching practices. For these reasons, we consider the observations of instructors when not using ITG to be a reasonable representation of teaching practices in undergraduate geoscience classrooms.

We use the instructional categories of Budd et al. (2013) to document differences in teaching practice because they are established in the literature and provide a convenient way to make comparisons. The RTOP was not developed or calibrated in a way that allows us to quantify the importance of score change in a linear way. The limitation to this approach is that scores that fall near the cutoffs for each category may differ from each other by only a few points, but are assigned to different instructional categories.

## Results

### Descriptive statistics

Observed instructors include 42.2% female and 57.8% male faculty, who reported that they are 33.2% full professors, 19.2% associate professors, 16.1% assistant professors, 22.4% lecturers, 8.4% adjunct professors, and 0.7% other (Fig. 2). Observations occurred at doctoral universities (47.4%), master’s colleges and universities (27.2%), baccalaureate colleges (5.9%), baccalaureate/associate’s colleges (1.0%), associate’s

colleges (18.1%), and one tribal college (0.3%), based on the Carnegie Classification System for Institutions of Higher Education (Carnegie, 2019). Classes observed were introductory level (64.3%), majors’ courses (35.3%), or considered “other” (0.3%), and include course enrollments classified as small ( $\leq 30$  students; 51.0%), medium (with 31–79 students; 30.4%), or large (with  $\geq 80$  students; 18.5%; Fig. 2).

RTOP scores for all 345 observations range from 13 to 89, with an average score of 41.9 (SD = 15.4; Table 4). There are 83 observations of teacher-centered classes (24.1% of observations), 163 observations of transitional classes (47.2%), and 99 observations of student-centered classes (28.7%), which includes all scores (some instructors are represented more than once; Fig. 4; Table 4). The average RTOP score when only the first ITG or first non-ITG observation is included (each instructor represented only once,  $n = 287$ ) is 40.5 (SD = 15.4). RTOP scores from the 50 observations of classrooms using ITG range from 34 to 75, and have an average score of 54.0 (SD = 10.5; Table 4). Thirty of the 50 ITG observations have scores in the student-centered instructional category (60%), 20 in the transitional instructional category (40%), and none in the teacher-centered category (Fig. 5). RTOP scores for the 295 non-ITG observations range from 13 to 89 with an average of 39.8 (SD = 15.2, Table 3). Of the 295 non-ITG observations, 69 are in the student-centered instructional category (23.4%), 142 are in the transitional instructional category (48.1%), and 84 are in the teacher-centered instructional category (28.5%; Fig. 5).

Five non-geoscience instructors in non-geoscience courses (philosophy, history, religious studies, biology, and nursing) were observed teaching with ITG materials, and had RTOP scores ranging from 43 to 75.

### Demographic and population comparisons

A multiple regression was run using all first ever RTOP observations ( $n = 287$ ) to predict RTOP score using the demographic characteristics of gender, institution type, instructor’s position, course type, and course size. While the model statistically significantly predicted RTOP score ( $F(5,279) = 3.428$ ,  $p = 0.005$ , adj.  $R^2 = 0.041$ ), only gender added significantly to the prediction ( $p = 0.001$ ), with females having predicted RTOP scores 6.0 points higher than males.

To assess how similar the populations of ITG and non-ITG users were, a Pearson’s chi-squared test was run for each demographic variable to test for independence between the two groups. There was a significant association between ITG use and institution type ( $p = 0.011$ , Cramer’s  $V = 0.228$ ), instructor

**Table 4** Summary of RTOP observation scores for all observations and for the subsets of ITG and non-ITG observations

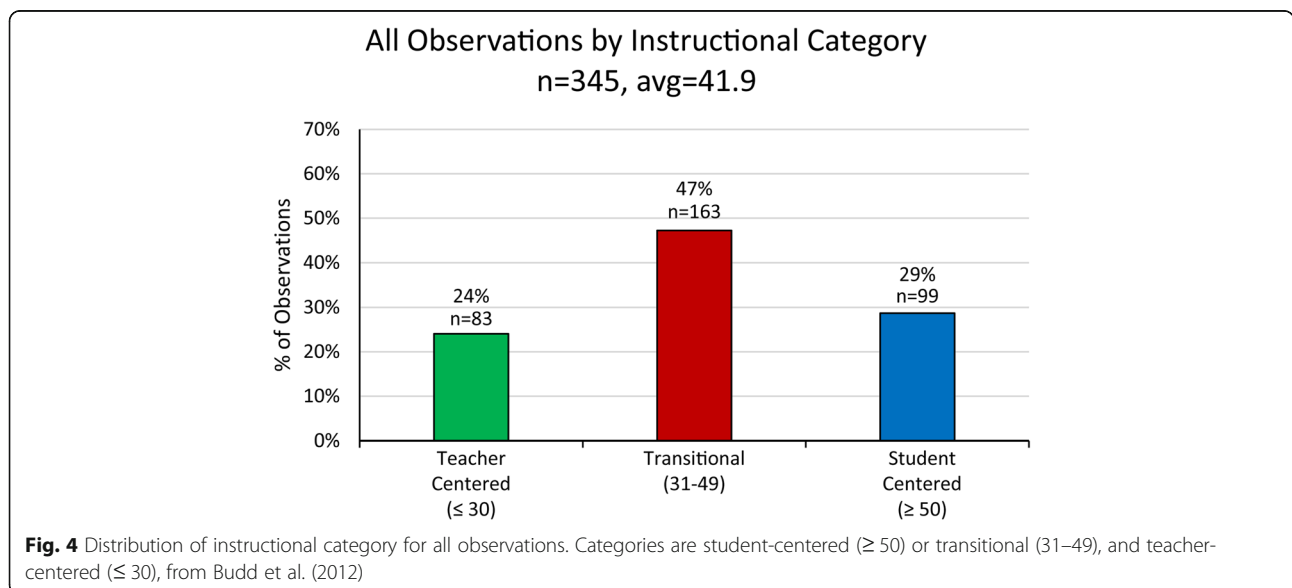
Mean	All Observations (n = 345)			Instructors with paired sets (n = 33)		
	41.9			NA		
	Score with ITG (n = 50)			Score with ITG (n = 33)		
With ITG average	54.0			54.0		
With ITG maximum	75.0			75.0		
With ITG minimum	34.0			34.0		
	Score w/o ITG (n = 295)			Score w/o ITG (n = 33)		
w/o ITG average	39.8			47.1		
w/o ITG maximum	89.0			74.0		
w/o ITG minimum	13.0			19.0		
	All observations by instructional category (with and w/o ITG)			Paired observations w/ ITG by instructional category		
Instructional category	# Observations (n = 345)	%	Average score by category	# Observations (n = 33)	%	Average score by category
Teacher centered	83	24.1	23.2	0	0	NA
Transitional	163	47.2	39.5	14	42.4	44.7
Student centered	99	28.7	61.6	19	57.6	60.8

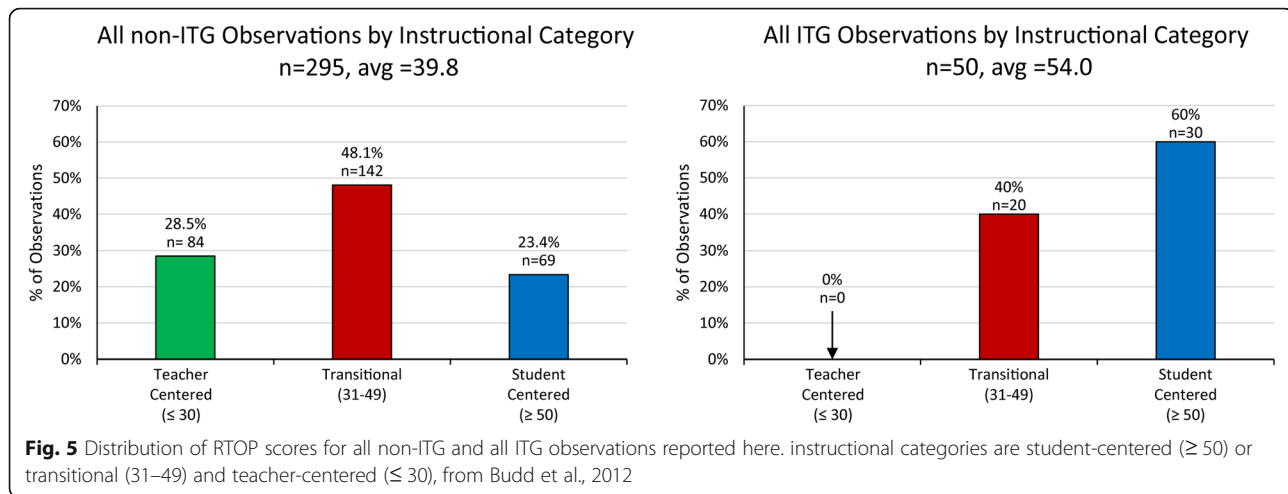
position ( $p = 0.001$ , Cramer’s  $V = 0.273$ ), and course type ( $p = 0.039$ , Cramer’s  $V = 0.151$ ). All associations were small (Cramer’s  $V < 0.3$ , Cohen, 1988). Post-hoc analysis using standardized residuals greater than 2 indicate that ITG users come from a higher count of Master’s universities, include more assistant professors, and taught more introductory courses. There was no significant association between ITG use and gender ( $p = 0.068$ ) or class size ( $p = 0.471$ ).

**What are the effects of teaching with ITG on the use of student-centered instructional strategies? (RQ1)**

Overall, 60% of instructors observed using ITG materials have RTOP scores in the student-centered

instructional category (score  $\geq 50$ ) and no instructors observed teaching with ITG have RTOP scores in the Teacher-Centered instructional category (score  $\leq 30$ ). A Welch’s unequal variances  $t$  test comparing ITG observations ( $n = 46$ ) and non-ITG observations ( $n = 241$ ) indicates that there is a significant difference ( $p < .001$ ) between RTOP scores of classes observed using ITG ( $M = 53.7$ ,  $SD = 10.4$ ) and classes observed not using ITG ( $M = 38.9$ ,  $SD = 15.4$ ). The effect size Hedge’s  $g$  is 1.01, which is considered large (Cohen, 1988). To test if the ITG users were representative of the non-ITG user’s population, an independent samples  $t$  test was run comparing the RTOP scores of paired instructors when not using ITG with





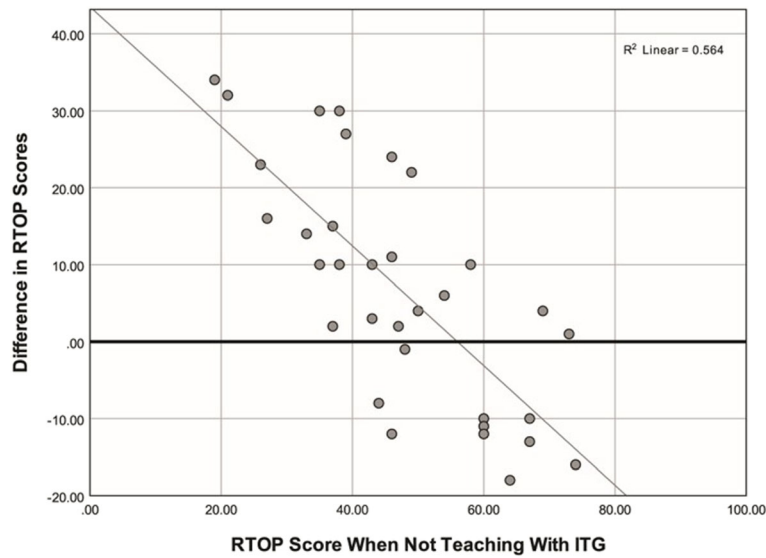
RTOP scores of the entire population when not using ITG. Because materials developers are known to have participated in intense PD associated with ITG (Egger et al., 2019) contributing to high non-ITG RTOP scores, we removed their scores from the population of paired instructors. When materials developers are not included, RTOP scores for the remaining paired instructors when not teaching with ITG (excluding materials developers,  $n = 23$ ,  $M = 41.7$ ,  $SD = 12.3$ ) are not significantly different ( $p = 0.391$ ) from those of unpaired instructors ( $n = 241$ ,  $M = 38.9$ ,  $SD = 15.4$ ). A Welch's unequal variances  $t$  test was run to compare the RTOP scores when teaching with and without ITG for this smaller sample of paired instructors that did not include the materials developers. This smaller sample of paired instructors had RTOP scores when teaching with ITG ( $n = 23$ ,  $M = 50.6$ ,  $SD = 9.5$ ) that were significantly higher than the RTOP scores of unpaired instructors teaching without ITG ( $n = 241$ ,  $M = 38.9$ ,  $SD = 15.4$ ,  $p < 0.001$ ) with a Hedge's  $g$  of 0.78, which is considered a medium effect size (Cohen, 1988).

A paired sample, two-tailed  $t$  test of the 33 paired observations indicates a significant difference ( $p = 0.013$ ) between paired instructors' RTOP scores when teaching with ITG ( $M = 54.0$ ,  $SD = 10.6$ ) and without ITG ( $M = 47.1$ ,  $SD = 14.7$ ). Cohen's  $d$  is 0.453, which is considered a small effect size (Cohen, 1988). There was no prescribed order of the paired observations (e.g., observed teaching with ITG before without ITG), so data do not reflect a progression from one type of teaching to another. There is no statistical difference in RTOP scores depending on whether an instructor was first observed using ITG or not.

To evaluate the impact that the use of ITG had on individual instructors during an individual class session, we compared each paired instructors' RTOP score when

teaching without ITG to the difference between their RTOP scores when teaching with and without ITG (Fig. 6). The difference between ITG and non-ITG scores (ITG RTOP score minus non-ITG RTOP score) ranges from  $-16$  to  $34$ , with an average difference of 8.5 for all paired observations (where a negative score indicates the RTOP score when not teaching with ITG is greater than the score when teaching with ITG for the same instructor). A Pearson's correlation indicates that there is a statistically significant, strong negative correlation between an instructor's RTOP score when teaching without ITG and the difference between their RTOP scores when teaching with and without ITG ( $r(31) = -0.751$ ,  $p < 0.001$ , Fig. 6). Normalizing the change (Marx & Cummings, 2007) to control for the number of RTOP points, an instructor could gain or lose still results in a significant, moderate negative correlation ( $r(31) = -0.647$ ,  $p < 0.001$ ). This means that instructors with the lowest RTOP scores for their non-ITG observations had the biggest change in RTOP score when they were observed using ITG. Instructors who are in the teacher-centered instructional category for non-ITG observations have the largest change in RTOP score (17–34 points increase; average change 25.6 points). Instructors in the transitional instructional category for non-ITG observations have differences between ITG and non-ITG RTOP scores that range from  $-12$  to  $30$  (average 11.4). Instructors in the student-centered instructional category for non-ITG observations have the lowest difference in RTOP scores, ranging from  $-16$  to  $6$  points (average =  $-4.8$  points).

Paired observations ( $n = 33$ ) indicate that all instructors who taught non-ITG classes in the teacher-centered ( $n = 4$ ) or transitional ( $n = 18$ ) instructional categories had RTOP scores that either were transitional or student-centered when using ITG (Fig. 7). Of these 22 instructors, 19 had higher RTOP scores when teaching with ITG, and three had lower RTOP scores when teaching with ITG, but were within the transitional

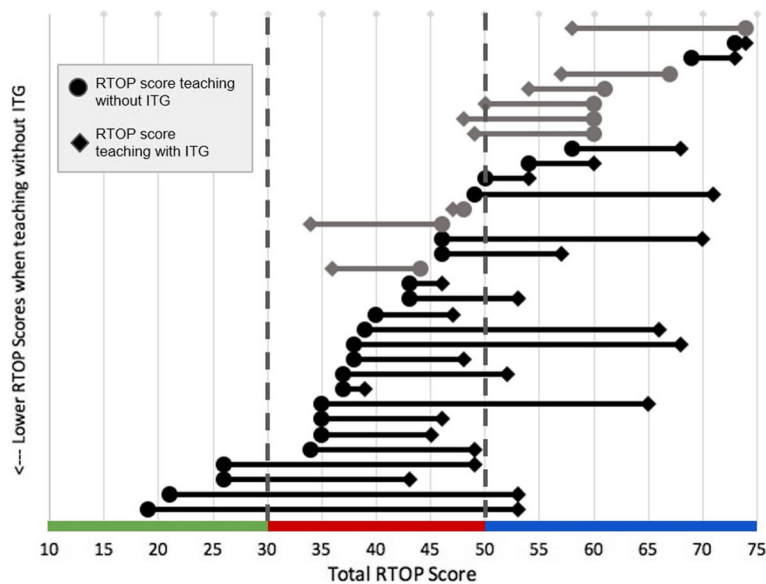


**Fig. 6** Difference in RTOP score for paired observations when teaching with and without ITG plotted against RTOP score when not teaching with ITG (positive difference indicates the ITG score is greater than the non-ITG score)

category for both observations. Two instructors with teacher-centered RTOP scores when teaching without ITG had student-centered RTOP scores when using ITG. Nine of the 11 instructors with non-ITG classes in the student-centered instructional category also taught student-centered ITG classes, and two instructors had RTOP scores in the transitional instructional category when they used ITG.

**RTOP subscale comparisons**

To determine if mean subscale scores in class periods observed using ITG were different than those not using ITG, an independent samples *t* test was used with a Bonferroni correction (appropriate for repeated tests across five subscales; in this case, a *p* value of less than .0125 is considered to be significant at  $\alpha = 0.05$ ). A significant difference was detected in four of the subscales:



**Fig. 7** Comparison of RTOP scores for individual instructors when teaching both with (diamond) and without (circle) ITG. Black lines connecting the two scores indicate RTOP scores when teaching with ITG were higher for that instructor; grey lines connecting the two scores indicate RTOP scores when teaching without ITG were higher for that instructor

lesson design and implementation, procedural knowledge, student-student interactions, and student-teacher relationships ( $p < 0.001$ ). No significant difference was detected for the propositional knowledge subscale based on whether the lesson was taught with or without ITG ( $p = 0.016$ ). The largest mean difference was found in the student-student interactions subscale (5.2), followed by student-teacher relationships (3.6), lesson design and implementation (2.4), and finally procedural knowledge (2.1). The mean difference of 0.9 points was non-significant for the propositional knowledge subscale. Thus, using ITG significantly impacts all subscales except that related to the accuracy and organization of content in the class session.

#### Does the level of teaching reform when using ITG depend on one's level of ITG participation? (RQ2)?

To determine whether a certain amount of professional development was needed to use ITG materials effectively, a two-way mixed ANOVA was used to compare the impacts of ITG use (within-subjects factor) across the three participation categories (between-subjects factor) for instructors with paired observations. There was not a statistically significant interaction between the use of ITG and the participation level on RTOP score ( $p = 0.715$ , Fig. 8), but both of the main effects were found to be statistically significant. The use of ITG led to a significant increase in RTOP scores ( $p = 0.016$ ) as did participation level ( $p = 0.001$ ). Pairwise comparisons reveal that the materials developers have significantly higher RTOP scores than both mentored instructors ( $p = 0.001$ ) and unmentored instructors ( $p = 0.024$ ) both when teaching with ITG and when teaching without

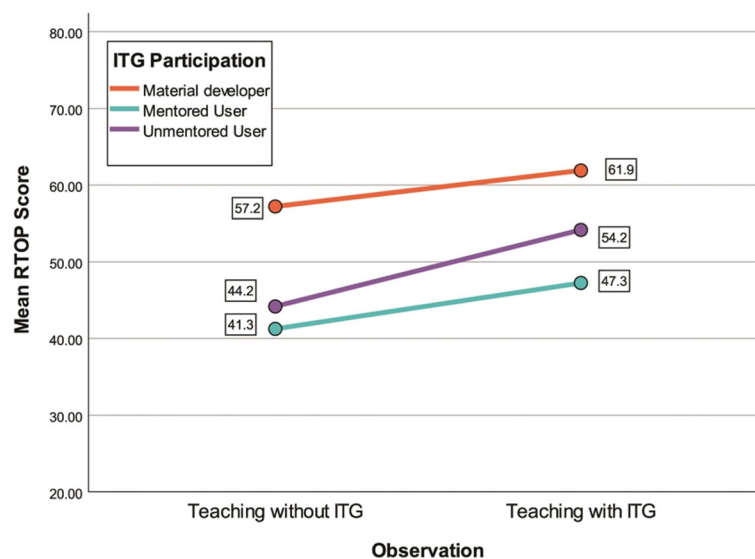
ITG. There was no significant difference between the RTOP scores of mentored and unmentored instructors ( $p = 0.502$ ). Thus, faculty participating at higher levels of engagement in creating ITG materials (materials developers) has higher RTOP scores with and without the use of ITG materials. Additionally, instructors at all levels of participation respond similarly to the use of ITG materials as evidenced by the significant main effect for RTOP score.

#### Do different instructors using the same ITG materials teach with similar levels of reform? (RQ3)

Six different ITG units were observed being taught by at least two different instructors. RTOP scores differed by between 2 and 13 points for different instructors teaching with the same ITG materials, and all observations with ITG materials have scores in the transitional and student-centered instructional categories (Fig. 9, Table 5). Thirteen of the seventeen instructors were also observed teaching without ITG, and the range of those RTOP scores has greater variability, including scores in the teacher-centered, transitional, and student-centered instructional categories. Instructors observed teaching with the same ITG unit has differences in their RTOP scores when teaching without ITG that range from 9 and 40 points (Table 5). Thus, similar levels of reform are observed for instructors teaching with the same ITG materials even though varying levels of reform are observed when the same instructors are teaching without ITG.

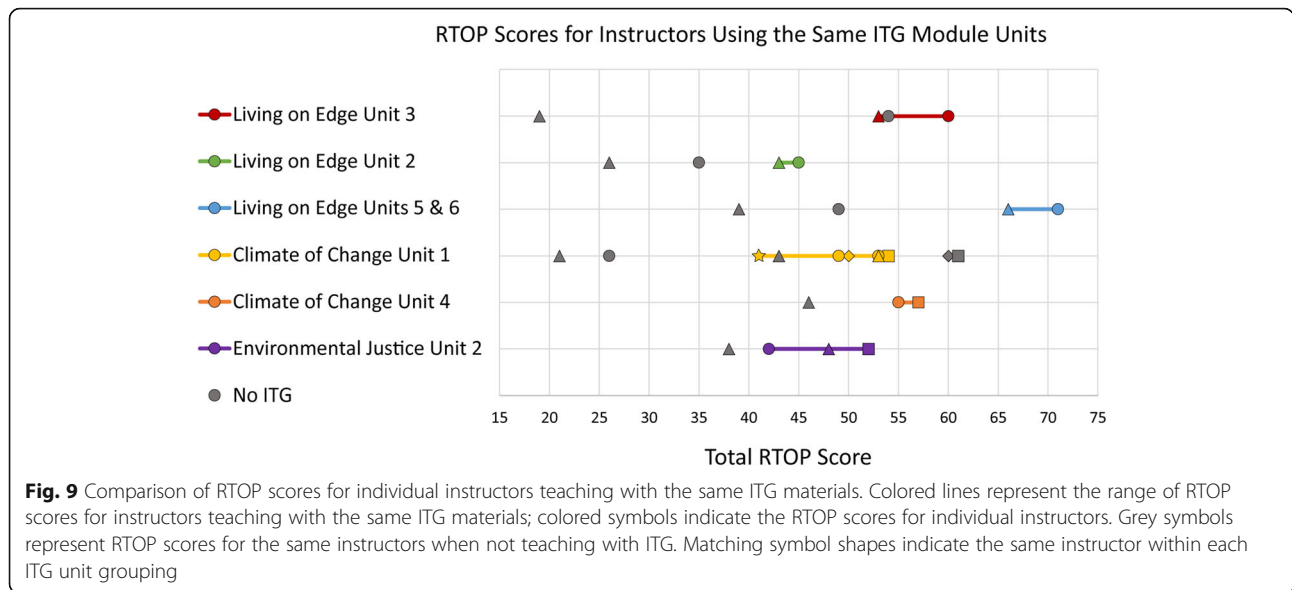
#### Discussion

Instructors who were observed for this project comprise a broadly representative distribution of geoscience



**Fig. 8** Mean RTOP scores for paired instructors' ITG and non-ITG observations, separated by instructors' participation in ITG Projects



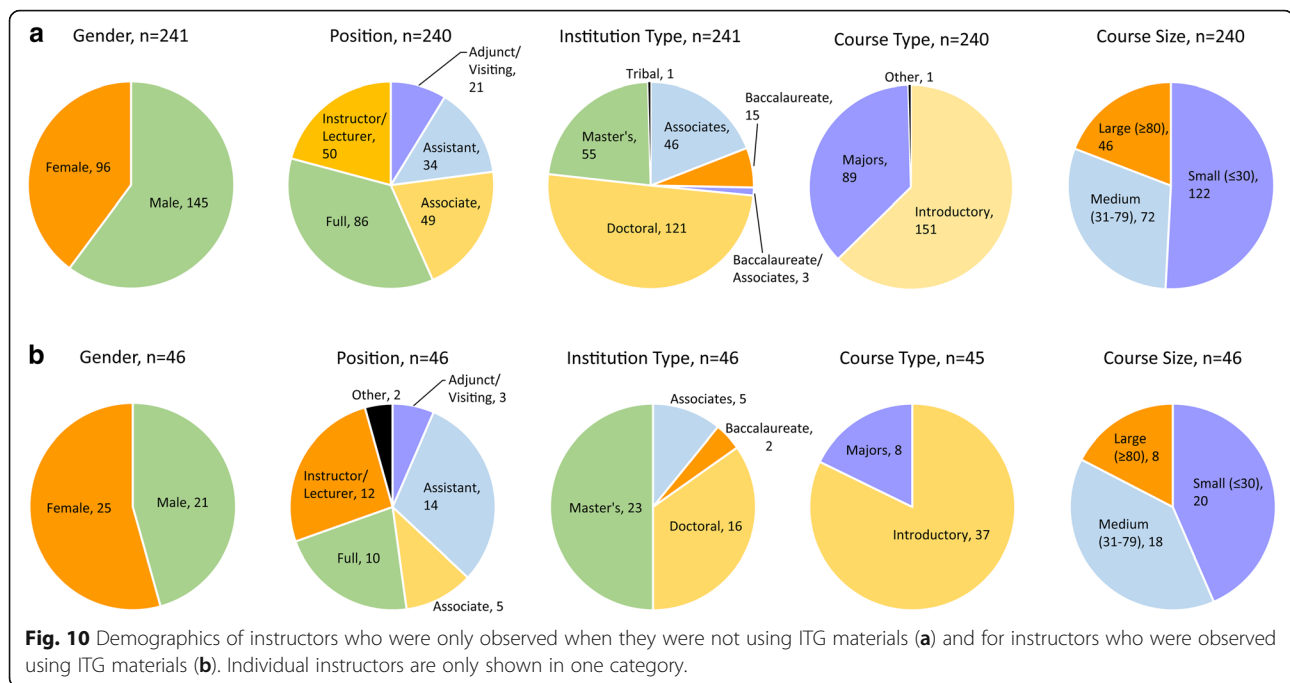


instructors in their institution types and positions (seniority), and reflect a range of geoscience course types and class sizes (Fig. 10). There are no correlations between scores and demographic factors of institution type, instructor position, course type, and course size. The demographic of gender does add significantly to the prediction of RTOP score. The sample of instructors who were observed using ITG materials differed from the

larger population, in that more taught at Master’s institutions, more were assistant professors, and more taught introductory courses. Since the ITG users did not significantly differ from the larger population based on gender, we posit that comparisons of their RTOP scores are influenced in the same ways as the general population, so comparisons of the two populations are not demographically influenced.

**Table 5** Comparison of RTOP scores for instructors teaching with the same ITG materials

ITG Module	RTOP score with ITG	RTOP score without ITG	Spread of RTOP scores with ITG	Spread of RTOP scores without ITG
Living on the Edge, Unit 3 $n = 2$	53	19	7	35
	60	54		
Living on the Edge, Unit 2 $n = 2$	45	35	2	9
	43	26		
Living on the Edge, Units 5 and 6 $n = 2$	66	39	5	10
	71	49		
Climate of Change, Unit 1 $n = 6$	50	60	13	40
	53	21		
	54	61		
	49	26		
	53	43		
Climate of Change, Unit 4 $n = 2$	41	–	2	–
	55	–		
Environmental Justice, Unit 2 $n = 3$	57	46	11	–
	42	–		
	53	–		
	48	38		



### InTeGrate materials support student-centered teaching practices (RQ1)

Given the structural design of ITG materials and the embedded student-centered pedagogies, our hypothesis is that teaching with ITG results in more reformed teaching than instruction without ITG. Comparisons of RTOP scores for all instructors teaching with ITG and all instructors teaching without ITG indicate that more reformed teaching practices (higher RTOP scores) are associated with the use of ITG materials. While instructors who use ITG modules may not reflect the general population of all instructors observed for this project, the comparison of RTOP scores for paired observations (individual instructors observed teaching both with and without ITG materials) also indicate that more reformed teaching practices (higher RTOP scores) are associated with ITG use. Thus, the correlation between using ITG materials and more student-centered teaching is true at the community level (comparison of overall ITG and non-ITG populations) and solidified at the individual level (comparison of paired observations).

Because the mean RTOP score of all ITG observations is significantly higher than the mean score of all non-ITG observations, we assert that the use of ITG materials is an important factor in the observed difference in levels of student-centered teaching. In an environment of change, ITG materials (teaching instructions, instructor stories, data, and assessments) serve as external resources instructors can access (external domain; Fig. 1, Table 1), which create a pathway to experimentation

with the new curriculum. This experimentation may be with ITG curriculum that provides opportunity to teach a new topic, use new pedagogical strategies, or both (domain of practice), in which increased RTOP scores provide evidence for a change sequence (a one-time change). RTOP observation scores provide a mechanism to measure instructors' domain of practice and while not intended to contribute to the instructional development of faculty, observed instructors could reflect on their RTOP scores to inform their teaching (personal domain, domain of consequence). Such reflection was not measured here.

The population of instructors with paired observations includes instructors whose RTOP scores when teaching without ITG span all instructional categories (Fig. 7), but none of these instructors were observed teaching teacher-centered classes (RTOP scores  $\leq 30$ ) when using ITG materials. Likewise, in the population of all instructors observed, there are no ITG observations in the teacher-centered instructional category. In terms of the change environment (Clarke & Hollingsworth, 2002; Fig. 1), our data show that when instructors use (enact) the external stimuli of ITG materials, their domain of practice changes (increased RTOP score) for that class session. Thus, ITG can play a role in supporting faculty efforts to teach transitional (RTOP scores 31–49) or student-centered class sessions (RTOP score  $\geq 50$ ). Eliminating less effective teacher-centered instruction and increasing the use of student-centered instruction is transformational, and indicates that ITG materials, or those following a similar set of principles and design

processes, can play an important role in promoting the increased use of student-centered instruction.

Among instructors who were observed teaching both with and without ITG, there is a strong negative correlation between their non-ITG RTOP score and the difference in the RTOP score when teaching with ITG (Fig. 6). This means that instructors who have lower RTOP scores when teaching without ITG show the greatest gains in RTOP score when teaching with ITG. This supports the idea that ITG materials are most beneficial in promoting reformed teaching for instructors whose teaching is teacher-centered or transitional when not using ITG. Thus, our data suggest that instructors, particularly those with little experience with student-centered teaching techniques, are likely to receive sufficient stimulus from the ITG materials (external domain) to drive a change sequence toward more student-centered class sessions when the materials are used.

RTOP subscale scores of instructors observed teaching with ITG were higher for all subscales except propositional knowledge. The propositional knowledge subscale encompasses what the teacher knows and how well they are able to organize and present the material in a learner-oriented setting (Budd et al., 2013). We interpret the lack of significant difference in this subscale as an indication that there is no difference in content knowledge between instructors who are or are not using ITG. The two subscales with the largest differences (classroom culture—communicative interactions and classroom culture—student-teacher relationships) are measures of the interactions among students or between students and the instructor. This aligns with the instructional strategies section of the ITG materials development rubric that includes explicit promotion of student-teacher and student-student interactions, which results in ITG materials that emphasize students working together and with their instructors. Similarly, it is likely that because ITG materials are aligned with the ITG materials development rubric, there are significant differences in the procedural knowledge and lesson design and implementation subscales. Individual RTOP items in these subscales relate directly to items in the resources and materials section of the ITG materials development rubric, and the supporting materials associated with ITG curriculum that guide ITG users to promote both student-student and student-instructor interactions. The rubric-aligned features of ITG materials, including the use of effective instructional strategies (e.g., instruction that promotes student engagement, metacognition, and communication of geoscience) result in active learning opportunities, problem solving, and scenarios that engage students in their own learning, which typify reformed teaching. Thus, the design of ITG materials is directly aligned with high scores in four of the five

RTOP subscales, which contributes to higher total RTOP scores when using ITG materials.

#### **Improved instruction does not depend on level of ITG participation (RQ2)**

Use of ITG materials can result in high RTOP scores, even for unmentored instructors who do not have extensive PD prior to their use (Figs. 5 and 8). In the environment of change, this means that as an external stimulus (external domain), the use of ITG materials provides a clear improvement in the domain of practice (more student-centered instruction). Thus, the effect of ITG materials on the domain of practice does not require additional stimuli such as extensive participation in the ITG project. The consistency with which ITG materials can be used without extensive PD means the ITG project can positively impact teaching beyond the duration of the project funding, even beyond the project-funded PD.

While ITG materials support the use of student-centered activities that instructors can easily access and use in class sessions, the embedded pedagogical strategies are not themselves limited to ITG materials. In a previous study, examination of comments made by RTOP observers for high scoring non-ITG classes resulted in three broad characteristics of teaching strategies used by instructors of student-centered classes: (1) instructors ask students questions; (2) students are engaged in meaningful conversations with other students; and (3) instructors assess student learning and adjust class sessions as appropriate (Teasdale, Viskupic, et al., 2017). While none of these are unique to ITG materials, the Instructional Strategies' section of the ITG materials development rubric is aligned with the observed characteristics.

Among the group comprising our 33 paired observations, nine instructors have RTOP scores for non-ITG observations that are higher than their scores when using ITG. Twenty of the fifty observations of instructors teaching with ITG have RTOP scores within the transitional, rather than student-centered, instructional category. This confirms that other factors in the environment of change can also play a role in whether using ITG materials (or similarly well-designed materials) will result in student-centered instruction.

While the ITG instructional materials on their own support more reformed teaching, other components of the ITG project (e.g., reading instructor stories, attending an ITG webinar, or workshop; Table 1) or other factors that are outside the scope of this project (e.g., conversations with colleagues, reading journal articles) may also be influential. Of the 46 instructors observed using ITG, 31 were part of an ITG-related PD program, either associated with materials development or

participation in an implementation project or research team (mentored instructors). The average difference between RTOP scores when teaching with and without ITG was 5.6 points for materials developers and 8.7 points for mentored instructors. The smaller difference for materials developers is consistent with the finding that student-centered instruction could be supported through high-quality professional development (Viskupic et al., 2019). In comparison, the average difference in RTOP score for unmentored instructors who were not part of an ITG-related project was 12.1 points. Of the factors we can measure (RTOP score, use of ITG, and participation in ITG-related PD), the difference in RTOP scores for unmentored instructors is apparently strongly impacted by the ITG materials. This is consistent with greater improvement of RTOP scores for instructors with lower scores for non-ITG observations (e.g., RQ2; Fig. 6) than by PD because these instructors had little or no ITG-related PD. In contrast, we cannot resolve the relative effects of ITG materials and ITG-related project PD for materials developers and mentored instructors. Unmentored instructors are more representative of the broad array of instructors who download ITG materials without benefiting from ITG related PD. As such, the differences between ITG and non-ITG observation RTOP scores of unmentored instructors indicate that even without ITG-related PD, instructors can reasonably expect that using ITG materials will increase the student-centeredness of that class session.

#### **Instructors using the same ITG materials teach with similar levels of reform (RQ3)**

Some ITG units were observed in use in multiple classrooms, giving us the opportunity to understand the range of implementation and its impact on RTOP score. Instructors who were observed teaching with the same ITG materials have vastly different RTOP scores when they are teaching without ITG, but have similar RTOP scores when they are teaching with the same ITG units (Table 5, Fig. 9). The relatively small variation in RTOP score for instructors teaching with the same ITG unit demonstrates that using the same ITG curriculum will result in similar levels of student-centered teaching regardless of the instructor.

This small range in RTOP scores for instructors teaching with the same ITG units (Fig. 9) was surprising considering that faculty may modify the curriculum to suit their own goals and unique teaching environments. Instructor materials for each ITG unit were designed to allow adaptability and ease of use so that materials could be modified to fit the instructor's needs. For example, instructors who teach shorter duration classes may have

cut out or condensed parts of the ITG activities to fit their allotted time. While classroom observers did not document the extent to which instructors modified ITG units from the published versions, observer comments indicate that many instructors adapted materials to varying extents. In the small samples examined for this study, we saw no correlation between RTOP score and either class size or class duration for instructors teaching with the same ITG unit. The strength of the reformed teaching strategies that are embedded in the ITG materials remain intact despite instructor modifications, as evidenced by similar RTOP scores when different instructors taught the same units.

The similarity in RTOP scores for individual ITG units persists among instructors with varying levels of participation in the ITG project. This supports the idea that ITG materials can be used consistently by different instructors, including those with minimal or no ITG-related professional development. While materials developers generally have higher RTOP scores than other instructors whether teaching with ITG or not, other work has shown that student learning and interest increases with use of ITG materials (Living on the Edge module) regardless of whether students are taught by the module author or not (Teasdale, Selkin, & Goodell, 2017).

The internally consistent use of ITG materials with respect to levels of student-centered instruction can be attributed to the common set of standards by which they were designed (e.g., Egger et al., 2019), which ensured they were aligned with the use of evidence-based instructional practices, and that those instructional practices were pervasive throughout each module (Steer et al., 2019). Additionally, as noted previously, ITG materials were developed by interdisciplinary teams of multiple authors plus an assessment team member and module editor, all of whom were from different institutions, thus providing diverse perspectives during the development of each module.

RTOP scores are consistent for instructors teaching with the same ITG unit but there are variations in RTOP score among the different units observed, even units within the same module. For example, RTOP scores for classes using ITG module Living on the Edge Units 5 and 6 (taught together in one class period) are approximately 20 points higher than RTOP scores for classes using ITG module Living on the Edge Unit 2 (Fig. 9). In addition to demonstrating the robustness of ITG materials, this observation suggests that some ITG units support student-centered teaching to a greater degree than others. This is not surprising, considering that the ITG materials development rubric was applied to ITG modules as a whole, and there is variation in the types of activities among the units within each module.

Variations in the degree to which different ITG units support student-centered teaching may explain why some instructors have lower RTOP scores when teaching with ITG materials than when teaching without ITG materials. Because RTOP scores for classes using specific ITG units are consistent and independent of the instructor, it is reasonable to expect that some instructors might have higher RTOP scores when not teaching with ITG if their typical teaching is student-centered.

Comparison of multiple instructors' use of the same materials also provides the opportunity to note that in the environment of change, the enactment of the same external supports (ITG materials) will have consistent effects on the instructors' domain of practice (RTOP scores). It is not yet clear if the small observed variations are related to other domains (personal domain or domain of consequence; Fig. 1), but our data show the ITG materials can result in the effective use of the embedded student-centered instructional strategies for multiple instructors with varying levels of previous experience with ITG and with varying instructional practices when not teaching with ITG.

#### **Mechanisms for developing a growth network for lasting change**

Our observation that unmentored instructors increase their use of student-centered practices when they use ITG materials indicates that adopting teaching materials with student-centered approaches is an effective way to modify class sessions for a typical instructor (e.g., without ITG-related PD). Such success raises the possibility that using ITG materials may increase an instructor's use of student-centered instruction in other courses or class sessions where ITG materials are not used. We attempted to test this idea by looking at RTOP scores for paired instructors in the context of the order that their ITG- and non-ITG observations occurred, but do not have a sufficient number of observations. A more focused effort to detect such transfer of student-centered instructional practices (e.g., transitioning from a limited change sequence to a growth network, Fig. 1) is addressed in the "Future work" section.

Other research has shown that broad changes to instructors' teaching practice can come from a focus on teaching and learning in PD (e.g., Czajka & McConnell, 2019; Derting et al., 2016). High RTOP scores of materials developers provide an example of the combined effect of intensive PD with the use of rigorously developed ITG materials containing embedded student-centered pedagogies, which may be a model for achieving the higher goal of reformed teaching. It is clear that PD plays an important role in promoting long term instructional change, and we have shown that materials adoption can also promote reformed teaching in a single

class period. However, it is unclear whether material adoption alone would aid in long term changes to teaching practices resulting in a growth network, or what combination of material adoption and additional PD would do so.

Any discussion of sustained change to one's teaching practice (growth network) must also recognize that such change requires instructors to also overcome potential barriers. Noteworthy barriers to the wholesale use of reformed teaching practices as part of a growth network includes access to training on the development of new course activities, time to create new materials and implement them, and lack of incentives (e.g., institutional or departmental support, e.g., Brownell & Tanner, 2012; Riihimaki & Viskupic, 2019). While ITG materials diminish or eliminate the barrier of time to develop lessons that use student-centered pedagogies, other barriers may impede a growth network for overall change. The question of whether ITG materials support change across an instructor's teaching practice required for a growth network is likely best resolved with a larger data set involving a strategically selected set of instructors, which falls outside the scope of the current project.

#### **ITG as a model for transforming STEM education**

Based on the significant increase in student-centered teaching and the absence of teacher-centered instruction observed in classes using ITG materials, it is clear that ITG materials can contribute to improving geoscience teaching. When coupled with quantified improvements in student learning and interest (e.g., Fortner et al., 2016; Teasdale, Selkin, & Goodell, 2017), we contend that ITG materials can have a positive impact in the "revolution" in disciplinary teaching that has been advocated for STEM disciplines (e.g., Kober, 2015; NASME, 2018; Olson & Riordan, 2012; Singer & Smith, 2013; Stains et al., 2018). The wide array of topics addressed in ITG modules (InTeGrate, 2019b) will likely suit many courses within and outside of the geosciences and can support increased use of student-centered instructional strategies.

Disciplinary transformations could be prompted by community-wide efforts similar to the ITG project, which by 2019 had reached more than 900 institutions in the USA and abroad (Gosselin, Manduca, et al., 2019). This represents a large proportion of all geoscience departments (reported by the American Geosciences Institute as 1940 globally and 963 in the USA; Wilson, 2018). Our data indicate that ITG materials can consistently result in reformed teaching when used, which suggests that other disciplines (or sub-disciplines) can use the ITG project model (see Gosselin, Manduca, et al., 2019) to provide similar external domain resources to faculty if important components of the ITG project are retained.



Consistent achievement of more student-centered teaching (a change sequence) with ITG is likely the result of robust curriculum development (Steer et al., 2019) and evaluation of materials (e.g., ITG assessment team and module editors; Egger et al., 2019) prior to publication. Similarly, it is likely that different faculty using the same ITG materials can teach with similar levels of reform (Fig. 9) because reformed teaching strategies are pervasive throughout the materials, and because ITG materials were developed for flexibility and adaptability by teams of module authors from diverse backgrounds (e.g., institution types). In addition to the large network of ITG materials developers using and promoting ITG materials, other ITG projects were specifically geared to broadly disseminate and implement ITG materials and offered professional development to support their adoption (e.g., Orr, C.H. & McDaris, J.R., 2019). ITG modules also include numerous external domain supports within the materials (teaching tips, instructor stories) that guide users through implementing a variety of student-centered teaching strategies, and provide guidance for each instructional strategy along with links to online resources (e.g., How to use a jigsaw activity in your classroom; Tewksbury, 2019). These features and activities helped combat the challenges of adoption widely encountered by other curriculum development projects (e.g. Henderson et al., 2011). The consistent (and significant) result of higher RTOP scores with the use of ITG materials suggests that this curriculum can be an important asset to instructors as a resource for the efficient implementation of student-centered teaching strategies in their classrooms. The use of student-centered teaching strategies and RTOP scores is more pronounced for instructors who use ITG and also participated in extensive PD (e.g., that of materials developers), which is ideal, but extensive PD is time-consuming and expensive, so not necessarily attainable across the geosciences in the short term. That said, unmentored instructors showed significant transformation toward student-centered teaching in single class sessions with the use of ITG. This shows that the “general user” can benefit significantly from use of carefully crafted community-derived curriculum. We see ITG materials as an important resource for increasing the level of student-centered instruction in geoscience classes, which along with continued efforts in discipline-based instructor PD (see Manduca et al., 2017), can transform geoscience teaching across the discipline. Other STEM disciplines with goals for teaching reform to support student learning can adapt the ITG Project model and build on its successes.

#### Future work

Research presented here examines instructors’ teaching *while* using ITG materials. Our comparisons of paired

ITG- and non-ITG observation scores show that the use of ITG can be transformative to the observed class session. This marks a change sequence, but our research design did not allow us to detect if using student-centered teaching when using ITG materials translates to other areas of instructors’ teaching practice as a growth network. Thus, our current data do not document if use of ITG materials can lead to a growth network with broader impacts on instructors’ teaching. In the context of the Interconnected Model of Professional Growth (Clarke & Hollingsworth, 2002), the use of an external resource as a single experiment in one’s teaching can be measured by RTOP data (domain of practice; Fig. 1). Future work will explore ways in which the experiment of using ITG might impact other areas of the change environment. For example, instructors may reflect on the outcomes of using ITG (domain of consequence), which may in turn, influence one’s beliefs (personal domain) and ultimately result in a growth network through changes to an instructor’s wholesale teaching practice (domain of practice; Table 1).

Future work will investigate the utility of ITG as a component of overall instructional change (growth networks) beyond the use of ITG (vs. our investigation here of ITG as a means of achieving student-centered instruction in a single class session). Multiple observations of instructors may help to more broadly characterize their teaching and examine non-ITG RTOP scores from observations made at intervals after ITG materials were first used. One hypothesis (not measured by this work) is that teaching with ITG facilitates instructors’ transfer of student-centered strategies throughout their teaching practice (a growth network), resulting in higher non-ITG scores than instructors who are observed prior to their experience with ITG.

Faculty interviews could also clarify details of how ITG influences an individual instructor’s pathway through the change environment (Clarke & Hollingsworth, 2002). One instructor’s personal domain may change with reflection from one use of ITG materials (external domain), while other instructors may need additional experimentation (e.g., with other ITG units or other curriculum that incorporates student-centered practices) to precipitate a change in their beliefs or attitudes, leading to a long-term change in practices. Once an instructor’s beliefs change to the point of wanting to modify their entire teaching practice, the interconnected model for the change environment could become a positive feedback loop where instructors are prompted to use more ITG (or similarly effective materials), resulting in a growth network, with changes to their entire teaching practice (Fig. 1).

#### Conclusions

For most instructors observed, RTOP observations reveal that more student-centered instructional practices occur

in class sessions in which ITG materials are used. While the largest change is observed for instructors with the least prior exposure to ITG project supports (e.g., programmatic ITG PD), ITG materials developers who had extensive ITG-related professional development have the highest RTOP scores of all groups examined. Thus, ITG materials are an effective external stimulus (external domain) in making change to one's domain of practice during class sessions when they are used. We contend that the use of community-authored ITG materials, which were developed following a consistent structure, are an effective complement to PD for instructors to increase their use of student-centered instructional practices.

The strength of the design process of ITG materials is evident from the overall increase in RTOP scores, and the absence of teacher-centered (RTOP scores  $\leq 30$ ) instruction when ITG materials are used. Even with the built-in adaptability of ITG materials, the embedded student-centered practices remain robust, as evidenced by very small variations in RTOP scores for different instructors using the same ITG materials.

Our data cannot determine whether use of ITG materials significantly impacts overall, long-term change to an instructor's teaching practice (e.g., a growth network), but does reveal that when in use, most instructors teach with more reformed student-centered practices than when not using ITG materials. A more transformational growth network with lasting change of an instructor's teaching practice may result from the use of ITG materials, which offer instructors the opportunity to experiment with student-centered curriculum. Future work is needed to address whether the use of ITG materials prompts instructors to reflect on student outcomes (learning, interest or interactions among students; domain of consequence) or changes instructors' beliefs or attitudes (personal domain). Such outcomes may lead to larger scale, long lasting change as a growth network (Clarke & Hollingsworth, 2002). Thus, the ITG materials have the potential to play an important role as instructors transform their teaching practice with a change sequence or even growth network.

Finally, the ITG project itself, which included multiple large-scale programs to create ITG curricular materials and to broadly disseminate and implement them, has widely impacted the geoscience community (Gosselin, Manduca, et al., 2019). Given the success reported here, particularly the complete absence of teacher-centered instruction when using ITG materials, the ITG project is an important component of transforming instruction. Previous concerns that the creation of materials posted online are not sufficient to support change to instructors' teaching (e.g., Henderson & Dancy, 2007) are not the case for the ITG materials. The ITG project model could be successfully

replicated by other disciplinary communities if important design elements are maintained, including community development teams and robust review and editing of materials. Ultimately, the development of a repository of high-quality curricular materials that employ student-centered teaching strategies are likely to have a long-term positive impact on reformed teaching, an important outcome that STEM communities continue to seek.

#### Abbreviations

AAAS: American Association for the Advancement of Science; ITG: InTeGrate materials; NASME: National Academy of Sciences, Engineering and Medicine; non-ITG: Not InTeGrate materials; NSF: National Science Foundation; PD: Professional development; RQ: Research question; RTOP: Reformed Teaching Observation Protocol; STEM: Science, Technology, Engineering and Mathematics; STEP: Science, Technology, Engineering, and Mathematics Talent Expansion Program

#### Acknowledgements

The authors wish to thank the team of observers who invested much time and effort in being trained to use the RTOP instrument, and in making the classroom observations for this project. We also thank the 287 instructors who agreed to have their classrooms observed.

#### Authors' contributions

RT, KR, KV, and DC made classroom observations, and contributed equally to the writing of the manuscript, including conducting data analysis and the creation of figures. All authors contributed substantially to the conceptualization of this project. All authors approved the final manuscript.

#### Funding

Funding for this work comes from the National Science Foundation (NSF) Division of Undergraduate Education (DUE) grant 1022844, and an NSF collaboration between the Directorates for Education and Human Resources (EHR) and Geosciences (GEO) under grant DUE - 1125331.

#### Availability of data and materials

The datasets generated during and/or analyzed during the current study are not publicly available due to the identifiable nature of the data.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Geological & Environmental Sciences, California State University, Chico, 400 W 1st St, Chico, CA 95929, USA. <sup>2</sup>School of the Earth, Ocean and Environment, University of South Carolina, 701 Sumter Street, EWS 617, Columbia, SC 29208, USA. <sup>3</sup>Department of Geosciences, Boise State University, 1910 University Drive, Boise, ID 83725-1535, USA. <sup>4</sup>Department of Earth Science, Utah Valley University, 800 W. University Parkway, Orem, UT 84058, USA. <sup>5</sup>Carleton College, Science Education Resource Center, Northfield, MN 55057, USA.

Received: 28 November 2019 Accepted: 17 September 2020

Published online: 28 October 2020

#### References

- Addy, T. M., & Blanchard, M. R. (2010). The problem with reform from the bottom up: instructional practises and teacher beliefs of graduate teaching assistants following a reform-minded university teacher certificate programme. *International Journal of Science Education*, 32(8), 1045–1071.
- American Association for the Advancement of Science (2011). *Vision and change: A call to action, final report*. D.C.: Washington.
- Amrein-Beardsley, A., & 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(1), 5–24.

- Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education*, 13(1), 1–12.
- Austin, A. E. (2011). Promoting evidence-based change in undergraduate science education. Presented at the Fourth committee meeting on status, contributions, and future directions of discipline-based education research, Washington, D.C.
- Benabentos, R., Hazari, Z., Stanford, J. S., Potvin, G., Marsteller, P., Thompson, K. V., ... Kramer, L. (2020). Measuring the implementation of student-centered teaching strategies in lower-and upper-division STEM courses. *Journal of Geoscience Education*, 1–15. <https://doi.org/10.1080/10899995.2020.1768005>.
- Bland, J. M., & Altman, D. G. (2003). Applying the right statistics: analyses of measurement studies. *Ultrasound in Obstetrics and Gynecology*, 22(1), 85–93.
- Bowling, B. V., Huether, C. A., Wang, L., Myers, M. F., Markle, G. C., Dean, G. E., ... Jacob, G. A. (2008). Genetic Literacy of Undergraduate Non-Science Majors and the Impact of Introductory Biology and Genetics Courses. *BioScience*, 58(7), 654–660.
- Bradforth, S. E., Miller, E. R., Dichtel, W. R., Leibovich, A. K., Feig, A. L., Martin, J. D., ... Smith, T. L. (2015). University learning: Improve undergraduate science education. *Nature*, 523(7560), 282–284.
- Brownell, S. E., & Tanner, K. D. (2012). Barriers to faculty pedagogical change: Lack of training, time, incentives, and... tensions with professional identity? *CBE—Life Sciences Education*, 11(4), 339–346.
- Budd, D. A., van der Hoeven Kraft, K. J., McConnell, D. A., & Vislova, T. (2013). Characterizing teaching in introductory geology courses: measuring classroom practices. *Journal of Geoscience Education*, 61(4), 461–475. <https://doi.org/10.5408/12-381.1>
- Carnegie Classification (2019). The Carnegie Classification of Institutions of Higher Education. Retrieved from <http://carnegieclassifications.iu.edu>
- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and teacher education*, 18(8), 947–967.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, (vol. 2). Hillsdale, NJ: Erlbaum.
- Czajka, C. D., & McConnell, D. (2019). The adoption of student-centered teaching materials as a professional development experience for college faculty. *International Journal of Science Education*, 41(5), 693–711.
- Derting, T. L., Ebert-May, D., Henkel, T. P., Maher, J. M., Arnold, B., & Passmore, H. A. (2016). Assessing faculty professional development in STEM higher education: Sustainability of outcomes. *Science Advances*, 2(3), e1501422–e1501422.
- Dunn, A., MacKay, R., & Resor, P. (2019). Earth's Thermostat InTeGrate Module. [https://serc.carleton.edu/integrate/teaching\\_materials/global\\_energy/index.html](https://serc.carleton.edu/integrate/teaching_materials/global_energy/index.html).
- Ebert-May, D., Derting, T. L., Henkel, T. P., Middlemis Maher, J., Momsen, J. L., Arnold, B., & Passmore, H. A. (2015). Breaking the cycle: future faculty begin teaching with learner-centered strategies after professional development. *CBE—Life Sciences Education*, 14(2), ar22.
- Eddy, S. L., Converse, M., & 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*, 14(2), ar23.
- Egger, A. E., Bruckner, M. Z., Birnbaum, S. J., & Gilbert, L. A. (2019). Facilitating the development of effective interdisciplinary curricular materials. In D. C. Gosselin, A. E. Egger, & J. J. Taber (Eds.), *Interdisciplinary Teaching About Earth and the Environment for a Sustainable Future* (pp. 45–68). Springer, Cham. <https://doi.org/10.1007/978-3-030-03273-9>.
- Fairweather, J. (2008). *Linking evidence and promising practices in science, technology, engineering, and mathematics (STEM) undergraduate education*. Washington, DC: Board of Science Education, National Research Council, The National Academies.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafo, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415.
- Gosselin, D. C., Egger, A. E., & Taber, J. J. (Eds.). (2019). *Interdisciplinary teaching about Earth and the environment for a sustainable future*. Springer.
- Gosselin, D. C., Manduca, C. A., Bralower, T., & Egger, A. E. (2019). Preparing students to address societally relevant challenges in the geosciences: The InTeGrate approach. In D. C. Gosselin, A. E. Egger, & J. J. Taber (Eds.), *Interdisciplinary teaching about Earth and the environment for a sustainable future* (pp. 25–44). Springer, Cham. <https://doi.org/10.1007/978-3-030-03273-9>
- 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*, 66(1), 64–74.
- Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: an analytic review of the literature. *Journal of Research in Science Teaching*, 48(8), 952–984.
- Henderson, C., & Dancy, M. H. (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*, 3(2), 020102.
- Henderson, C., Finkelstein, N., & Beach, A. (2010). Beyond dissemination in college science teaching: An introduction to four core change strategies. *Journal of College Science Teaching*, 39(5), 18.
- Henderson, C., Mestre, J. P., & Slakey, L. L. (2015). Cognitive science research can improve undergraduate STEM instruction: What are the barriers? *Policy Insights from the Behavioral and Brain Sciences*, 2(1), 51–60.
- 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*, 99(5), 783–818.
- InTeGrate (2019a). About the InTeGrate Project. <https://serc.carleton.edu/integrate/about/index.html>.
- InTeGrate (2019b). InTeGrate Teaching Materials. [https://serc.carleton.edu/integrate/teaching\\_materials/index.html](https://serc.carleton.edu/integrate/teaching_materials/index.html).
- Iverson, E. R., & Wetzstein, L. (2020). Connecting learning about the earth to societal issues: Downstream effects on faculty teaching. In J. Ostrow (Ed.), *Teaching about Sustainability across Higher Education Coursework, New Directions for Teaching and Learning*, No. 161. San Francisco: Jossey-Bass.
- Jang, H., Kim, E.-J., & Reeve, J. (2012). Longitudinal test of self-determination theory's motivation mediation model in a naturally-occurring classroom context. *Journal of Educational Psychology*, 104, 1175–1188. <https://doi.org/10.1037/a0028089>.
- Kastens, K. A., & Manduca, C. A. (2017). Using systems thinking in the design, implementation, and evaluation of complex educational innovations, with examples from the InTeGrate project. *Journal of Geoscience Education*, 65(3), 219–230.
- Kastens, Kim. (2019). InTeGrate Final Evaluation Report, [https://d320qogomya1dw8.cloudfront.net/files/integrate/about/itg\\_finalevalreport.pdf](https://d320qogomya1dw8.cloudfront.net/files/integrate/about/itg_finalevalreport.pdf)
- Keith, C. J. (2019). Faculty members' use of learner-centered instruction at institutions in the United States. *International Journal of Teaching and Learning in Higher Education*, 31(3), 378–389.
- Kober, N. (2015). *Reaching Students: What Research Says About Effective Instruction in Undergraduate Science and Engineering*. In Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Kuh, G. D. (2008). Excerpt from high-impact educational practices: What they are, who has access to them, and why they matter. *Association of American Colleges and Universities*, 19–34.
- Lawson, A., Benford, R., Bloom, I., & Carlson, M. (2002). Evaluating college science and mathematics instruction. *Journal of College Science Teaching*, 31(6), 388–393.
- Lund, T. J., Pilarz, M., Velasco, J. B., Chakraverty, D., Rosploch, K., Undersander, M., & Stains, M. (2015). The best of both worlds: Building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. *CBE—Life Sciences Education*, 14(2), ar18.
- Macdonald, R. H., Beane, R. J., Baer, E. M., Eddy, P. L., Emerson, N. R., Hodder, J., ... Ormand, C. J. (2019). Accelerating change: The power of faculty change agents to promote diversity and inclusive teaching practices. *Journal of Geoscience Education*, 67(4), 330–339.
- MacIsaac, D., & Falconer, K. (2002). Reforming physics instruction via RTOP. *The Physics Teacher*, 40(8), 479.
- Manduca, C. A., Iverson, E. R., Luxenberg, M., Macdonald, R. H., McConnell, D. A., Mogk, D. W., & Tewksbury, B. J. (2017). Improving undergraduate STEM education: The efficacy of discipline-based professional development. *Science Advances*, 3(2), e1600193. <http://doi.org/10.1126/sciadv.1600193>
- Marbach-Ad, G. C., Egan, L. C., & Thompson, K. V. (2015). A disciplinary teaching and learning center: Applying pedagogical content knowledge to faculty development. In G. C. Weaver, W. D. Burgess, A. L. Childress, & L. Slakey (Eds.), *Transforming Institutions: Undergraduate STEM Education for the 21st Century*, (pp. 297–311). Purdue University Press.

- Markley, C. T., Miller, H., Kneeshaw, T., & Herbert, B. E. (2009). The relationship between instructors' conceptions of geoscience learning and classroom practice at a research university. *Journal of Geoscience Education*, 57(4), 264–274.
- Marshall, J. C., Smart, J., Lotter, C., & 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(6), 306–315.
- Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, 75(1), 87–91.
- McConnell, D. A., Steer, D. N., & Owens, K. D. (2003). Assessment and active learning strategies for introductory geology courses. *Journal of Geoscience Education*, 51(2), 205–216.
- Multon, K. D. (2010). Interrater reliability. In D. M. Nougherty, & B. Frey (Eds.), *Encyclopedia of Research Design*, (pp. 626–628). Thousand Oaks, California: Sage.
- National Academies of Sciences, Engineering, and Medicine. (2018). Indicators for monitoring undergraduate STEM education. Washington, DC: The National Academies Press. <https://doi.org/https://doi.org/10.17226/24943>.
- National Research Council (2012). Discipline-based education research: Understanding and improving learning in undergraduate science and engineering. (S. Singer, N. R. Nielsen, & H. A. Schweingruber, Eds.). Washington, D.C.: The National Academies Press. <https://doi.org/https://doi.org/10.17226/13362>
- O'Connell, K., Bruckner, M. Z., Manduca, C. A., & Gosselin, D. C. (2016). Supporting interdisciplinary teaching about the Earth with the InTeGrate website. *Journal of Environmental Studies and Sciences*, 6(2), 354–359.
- Olson, S., & Riordan, D. G. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Report to the President. Executive Office of the President.
- Orr, C., H., & McDaris, J. R. (2019). Supporting implementation of program-level changes to increase learning about earth. In D. C. Gosselin, A. E. Egger, & J. J. Taber (Eds.), *Interdisciplinary Teaching About Earth and the Environment for a Sustainable Future* (pp. 69–89). Springer, Cham. <https://doi.org/https://doi.org/10.1007/978-3-030-03273-9>
- Piburn, M., & Sawada, D. (2001). Reformed Teaching Observation Protocol (RTOP) Reference Manual, ACEPT Technical Report No 1N00-3, 1–44
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of engineering education*, 93(3), 223–231.
- Riihimaki, C. A., & Viskupic, K. (2019). Motivators and inhibitors to change: Why and how geoscience faculty modify their course content and teaching methods. *Journal of Geoscience Education*, 68(2), 115–132. <https://doi.org/10.1080/10899995.2019.1628590>.
- Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics*, 102(6), 245–253.
- Singer, S., & Smith, K. A. (2013). Discipline-based education research: Understanding and improving learning in undergraduate science and engineering. *Journal of Engineering Education*, 102(4), 468–471.
- Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE Life Sciences Education*, 12(4), 618–627.
- Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., ... Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science*, 359(6383), 1468–1470.
- Steer, D., Iverson, E. R., Egger, A. E., Kastens, K. A., Manduca, C. A., & McConnell, D. (2019). The InTeGrate materials development rubric: a framework and process for developing curricular materials that meet ambitious goals. In D. C. Gosselin, A. E. Egger, & J. J. Taber (Eds.), *Interdisciplinary Teaching About Earth and the Environment for a Sustainable Future* (pp. 25–43). Springer, Cham. <https://doi.org/https://doi.org/10.1007/978-3-030-03273-9>.
- Stes, A., Min-Leliveld, M., Gijbels, D., & Van Petegem, P. (2010). The impact of instructional development in higher education: The state-of-the-art of the research. *Educational Research Review*, 5(1), 25–49.
- Sunal, D. W., Hodges, J., Sunal, C. S., Whitaker, K. W., Freeman, L. M., Edwards, L., ... Odell, M. (2001). Teaching science in higher education: Faculty professional development and barriers to change. *School Science and Mathematics*, 101(5), 246–257.
- Teasdale, R., Selkin, P., & Goodell, L. (2017). Evaluation of student learning, self-efficacy and perception of the value of geologic monitoring from Living on the Edge, an InTeGrate curriculum module. *Journal of Geoscience Education*, 66(3), 186–204.
- Teasdale, R., Viskupic, K., Bartley, J. K., McConnell, D., Manduca, C., Bruckner, M., Farthing, D., Iverson, E. (2017). A multidimensional assessment of reformed teaching practice in geoscience classrooms. *Geosphere*, 13(2), 608–627. <http://doi.org/https://doi.org/10.1130/GES01479.1>.
- Tewksbury, B. (2019). Pedagogy in Action: Jigsaws. <https://serc.carleton.edu/sp/library/jigsaws/index.html>
- Umbach, P. D., & Wawrzynski, M. R. (2005). Faculty do matter: The role of college faculty in student learning and engagement. *Research in Higher Education*, 46(2), 153–184.
- Viskupic, K., Ryker, K., Teasdale, R., Manduca, C., Iverson, E., Farthing, D., Bruckner, M. Z., & McFadden, R. (2019). Classroom Observations Indicate the Positive Impacts of Discipline-Based Professional Development. *Journal for STEM Education Research*, 1–28.
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, 23(3), 34–41.
- Wieman, C., & Gilbert, S. (2014). The teaching practices inventory: A new tool for characterizing college and university teaching in mathematics and science. *CBE—Life Sciences Education*, 13(3), 552–569.
- Wieman, C., Perkins, K., & Gilbert, S. (2010). Transforming science education at large research universities: A case study in progress. *Change: The Magazine of Higher Learning*, 42(2), 6–14.
- Wilson, C. (2018). *2018 Directory of Geoscience Departments*. American Geosciences Institute, (53rd ed., ). VA: Alexandria.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)