RESEARCH   **Open Access**

Exploring STEM postsecondary instructors' accounts of instructional planning and revisions

Robert Erdmann[1], Kathryn Miller[2] and Marilyne Stains[3*]

Abstract

Background: Local and national initiatives to improve the learning experiences of students enrolled in Science, Technology, Engineering, and Mathematics (STEM) courses have been on-going for a couple of decades with a heightened momentum within the last 10 years. However, recent large-scale studies have demonstrated that transmission of information is still the primary mode of instruction in STEM courses across the undergraduate curriculum. The limited impact of instructional change reform efforts can be partly explained by the one-sided focus of educational research on the development of evidence-based instructional practices and production of evidence demonstrating their impact on student learning. This has been done at the expense of understanding faculty members' instructional practices and mindsets about teaching and learning that underlie their practices. This study addresses this gap in the literature by characterizing STEM instructors' instructional intentions and reflections on their teaching performance for a week of instruction. Data was collected through semi-structured interviews with 42 STEM faculty members from one doctorate-granting institution in the USA.

Results: STEM instructors in this study had teacher-centric mindsets with respect to their instructional planning (e.g., content-focused learning goals, lecture is seen as an engagement strategy). We found that these instructors mostly saw formative assessment tools as engagement strategy rather than tools to monitor student learning. Reflections on their level of satisfaction with their week of teaching focused heavily on content coverage and personal feelings and minimally considered student learning. Finally, we found that pedagogical discontent was not a driver of planned course revisions.

Conclusions: This study identifies mismatches between STEM instructors' teaching mindsets and current approaches to instructional change. STEM instructors in this study paid minimal attention to student learning when considering course-level revisions and many of their reflections were anchored in their personal feelings. However, instructional reform strategies often attempt to convince faculty of a new approach by demonstrating its impact on student learning. The misalignment identified in this study further highlights the need to better characterize STEM instructors' cognition around teaching so that reform efforts can better meet them where they are.

Keywords: Instructional planning, Pedagogical discontent, Faculty cognition, Professional development, Postsecondary

* Correspondence: mstains@virginia.edu
[3]Department of Chemistry, University of Virginia, Charlottesville, VA
22904-4319, USA
Full list of author information is available at the end of the article

Introduction

The recruitment of the next generation of scientists and science teachers as well as the preparation of a science, technology, engineering, and mathematics (STEM) literate populace rests, in part, upon the learning environments students experience in postsecondary science classrooms. Although recent studies indicate that undergraduate student attrition in STEM majors, at 48%, is lower than other non-STEM fields (Mervis, 2014), instructional changes in STEM courses could reduce that rate (American Association for Higher Education, 2000; National Research Council, 1999, 2003b, 2011, 2012; Olson & Riordan, 2012; Project Kaleidoscope, 2002, 2006; Seymour & Hewitt, 1997). National initiatives to improve undergraduate STEM education (e.g., National Science Foundation, 2013; Association of American Universities, 2017) build upon decades of research and development of effective instructional practices (National Research Council, 1999, 2003b, 2011, 2012; Project Kaleidoscope, 2002, 2006). These evidence-based instructional practices (EBIPs) have been demonstrated to promote students' conceptual understanding and attitudes toward STEM (Freeman et al., 2014; Handelsman et al., 2004; National Research Council, 2011, 2012), with the greatest impacts observed among women and members of under-represented groups (Olson and Riordan, 2012). By improving retention rates and attracting a more diverse student population to the sciences, these practices can help address the national need to educate STEM-literate graduates and advance workforce development goals. The challenge is fostering their effective use on a national scale.

While, for decades, efforts were focused on the research and development of these EBIPs, characterizing the extent of their use by STEM faculty are recent endeavors. Within the last decade, several national survey studies have been conducted in order to capture faculty's level of awareness and implementation of EBIPs within a specific STEM discipline (Apkarian & Kirin, 2017; Borrego et al., 2010; Eagan, 2016; Henderson & Dancy, 2009; Macdonald et al., 2005; Prince et al., 2013; Zieffler et al., 2012). These surveys reveal that, while STEM faculty are aware of EBIPs, the majority do not implement these practices. Recently, two large-scale studies focused on characterizing instructional practices in STEM courses through classroom observations have corroborated this limited uptake and the predominance of lecturing (Stains et al., 2018; Teasdale et al., 2017).

The lack of integration of EBIPs into STEM courses exists for several reasons. First, STEM faculty lack experiences that would familiarize them with the functioning of these practices. Indeed, they typically receive little pedagogical training during their graduate and postdoctoral work and, for the most part, did not experience these practices as students. Second, the promotion strategies employed by developers of EBIPs have significant drawbacks (Borrego & Henderson, 2014; Henderson et al., 2011; Penberthy &

Millar, 2002). Third, the climate, culture, and rewards structure of STEM departments at research-intensive institutions significantly influence faculty's decisions to change their instructional practices (Henderson & Dancy, 2007; Lund and Stains, 2015; Shadle et al., 2017). Finally, studies have demonstrated that even when faculty adopt EBIPs, they often adapt them, sometimes at the peril of the critical elements that make the practice effective (Henderson & Dancy, 2009; Turpen & Finkelstein, 2009). This gap between developers' vision and actual implementation by faculty can be partly explained by the one-sided focus of educational research on the development of EBIPs and production of evidence demonstrating their impact on student learning. This has been done at the expense of understanding faculty's day-to-day thought processes when planning instruction and instructing outside the context of a reform effort (National Research Council, 2012; Talanquer, 2014). Characterizing these thought processes can enhance the design of professional development programs to meet faculty where they are. The main goal of this study is to provide insight into these day-to-day thought processes by exploring the following research question: How do STEM faculty plan the teaching of a week of content and how do they reflect on this experience?

Instructional thinking

The discipline-based education research (DBER) community in the USA has recently developed a better understanding of the instructional practices enacted in STEM undergraduate classrooms at the postsecondary level. However, the characterization of instructors' thinking has been mostly overlooked in DBER despite its importance in an era of instructional change reform efforts:

The value of this research for teacher educators and prospective teachers is that it provides insights into the mental lives of teachers. Those insights can be used to challenge preservice teachers' thinking and to expand their view of the teaching act. (Clark & Lampert, 1986, p. 27).

Clark and Peterson (1986) identified three types of thought processes, all unobservable: (a) instructors' planning (before and after instruction), (b) instructors' interactive thoughts and decisions during instruction, and (c) instructors' theories and beliefs. The literature points to a clear distinction between the thought processes that take place during teaching versus those that occur outside the teaching event (Clark & Peterson, 1986; Gess-Newsome, 2015). The latter type is the focus of this study. Instructional planning refers to both the thought processes that instructors engage in when they design a class session as well as the reflections that occur after classroom instruction. Instructional planning is viewed in the literature as a mental process—"a set of basic psychological processes in which a teacher visualizes the future, inventories means and ends, and constructs a framework to guide his or her future

action" (Clark & Peterson, 1986, p. 260) and a set of activities, i.e., actions the instructors take when they plan a lesson (Clark & Peterson, 1986).

Research on instructional planning, which has mostly focused on secondary teachers, has found that teachers attend to various domains when planning instruction (So, 1997): objectives, content, activities, students, teaching approach, evaluation, and theories/beliefs. Notably, the literature is in agreement that secondary teachers rarely follow the prescribed, sequential process first advanced by Tyler (1950): (a) identify learning outcomes, (b) choose instructional activities, (c) sequence learning activities, and (d) design assessment tools and plan. Interestingly, elements of this process are still at the core of many modern instructional design approaches (e.g., backward design; Wiggins & McTighe, 1998) and are emphasized in the training of STEM instructors (e.g., Scientific Teaching; Handelsman et al., 2004). Due to the ubiquity of these elements in the literature and pedagogical training approaches of STEM instructors, this study specifically explored STEM faculty's thinking about the (1) learning goals, (2) assessment tools to measure students' progress and acquisition of these learning goals, (3) learning experiences to support students' achievement of these learning goals, and (4) the instructors' reflections on their instructional implementation and their planned revisions. These characteristics have been investigated primarily in isolation in the literature on STEM instructors at the postsecondary level. Below we provide a summary of findings for each of these elements.

Learning goals identification
While planning their courses, instructors are influenced by both internal and external factors, ranging from personal pedagogical history to expectations set by accreditation bodies (Geis, 1996). Drawing on interviews with dozens of instructors, Stark et al. (1988) highlighted the role the discipline of the instructor plays in directing course planning activities. In subsequent work, Stark (2000) found that learning goals fall into shared patterns within disciplines. For example, biology instructors focused primarily on knowledge acquisition goals, while mathematics instructors tended to mix knowledge acquisition and skill development goals. Textbooks are one way in which norms of a discipline are transmitted to instructors, as they provide a superstructure upon which instructors can scaffold their courses (Davila & Talanquer, 2010). Outside of the influence of the discipline, external bodies, institutional accreditation or otherwise, can play an important role in setting the direction of learning outcomes across the entire range of university subjects (Slevin, 2001; Bretz, 2012).

Little has been reported about the cognitive level of learning goals that instructors construct for their classes. Momsen et al. (2010) analyzed syllabi from a range of 100- and 200-level college biology courses, finding that almost 70% of learning goals described in these documents were at the knowledge or comprehension levels of Bloom's taxonomy (a hierarchy of learning goals based on cognitive demand, ranging from memorization of defined content to drawing conclusions and making judgement; Bloom et al., 1956). This result indicates that many introductory biology instructors frame their learning goals around the memorization and understanding of factual information. Further analysis revealed that assessments associated with the analyzed syllabi were also heavily focused on these lower-level cognitive skills.

Learning goals assessment
The decision-making process that STEM instructors employ when developing and/or selecting assessment tools is largely unstudied. Some insight can be gained from investigations at the secondary level. For example, while studying the reasoning of secondary science teachers planning assessment tasks, Tomanek et al. (2008) proposed a conceptual framework of factors that influence the process. Factors influencing task evaluation and selection that were highly cited by the science teachers included factors related to the task itself (such as the extent to which the task is effective in illustrating the level of student understanding), factors related to student ability levels (such as literacy and numeracy levels), and factors related to the curriculum (such as the level of task alignment with the overall curriculum progression).

Within the higher education classroom, "tests and term papers remain the most traditional and typical ways of assessing learning" (Menges & Austin, 2001, p. 1146). National reports have advocated a rethinking of the prominence of summative assessment, emphasizing the benefits of formative assessment within the college STEM classroom (National Research Council, 2003a). Survey-based studies in chemistry indicate that instructors have at least a mid-level understanding of the difference between these two types of assessment (Emenike, Raker, & Holme, 2013; Raker & Holme, 2014).

Engaging students
The planning of learning activities is necessarily constrained by the awareness of potential approaches and techniques. As a result, many studies of postsecondary instructor classroom practices simultaneously assay instructor knowledge of potential teaching techniques. In engineering, Borrego et al. (2010) observed that the adoption level of a teaching approach that an instructor was aware of ranged from 35% in the case of service learning to 87% in the case of student-active pedagogies, showing that many factors outside of instructor knowledge went into the instructor's decision to utilize a teaching technique. Empirical models indicate that classroom, departmental and institutional context, as well as instructor's beliefs about teaching and

learning constitutes some of the factors influencing peda-gogical decisions (Clarke & Hollingsworth, 2002; Gess-Newsome et al., 2003). A survey of geoscience instructors found that while more than half of respondents incorpo-rated interactive activities of some sort weekly or more fre-quently, this most commonly took the form of adding demonstrations or questions to the lecture, with instructors aware of, but not often using, other forms of engagement (Macdonald et al., 2005). A comparison of biology, chemis-try, and physics instructors within the same institution found that while there was only a minimal difference in EBIP awareness levels among the three disciplines, the adoption levels were above average in physics and below average in chemistry, further highlighting the influence of disciplinary context on an instructor's selection of learning activities (Lund and Stains, 2015).

Recent studies have begun to illuminate population-level patterns in learning experiences within the postsecondary STEM classroom. In geoscience, survey results were com-bined with classroom observation data to characterize over 200 courses (Teasdale et al., 2017). Within this broad sam-ple, 25% of classes were classified as student centered, 45% of classes were classified as transitional, and 30% of classes were classified as teacher centered. Interestingly, the degree of student vs. teacher centeredness was not correlated with any facet of instructor, classroom, or institutional demo-graphic information, arguing that classroom practices were more dependent on individual instructor pedagogical be-liefs and choices than external factors within this popula-tion. In mathematics, a study of the pre-calculus through calculus II sequence characterized nearly 900 courses of-fered by 223 different mathematics departments (Apkarian & Kirin, 2017). The frequency of primarily lecture-based in-struction increased through the curriculum progression, with 59% of pre-calculus courses, 65% of calculus I courses, and 74% of calculus II courses exhibiting primarily lecture. The results of these discipline-specific investigations align with a multi-disciplinary observational study of over 2000 class periods taken from a variety of STEM courses (Stains et al., 2018). Analysis of instructor and student behaviors led to the grouping of classrooms into three categories: di-dactic, where greater than 80% of class time included lec-ture; interactive lecture, where lecture was complemented with a sizeable proportion of student-centered activities; and student-centered, where student centered instruction made up a large percentage of class time. Fifty-five percent of class periods were classified as didactic, 27% of class pe-riods were classified as interactive lecture, and the remaining 18% were classified as student-centered. Disci-pline and classification were correlated, with mathematics and geology class periods more likely to be student-centered, biology class periods more likely to be interactive, and chemistry class periods more likely to be didactic than would be expected from the overall population.

Reflections-on-action: pedagogical satisfaction

Reflections-on-action, an idea first introduced by Schön (1983), refers to reflections about how teaching can be developed, changed, or improved after the teaching event has taken place. Research at the K-12 level has demonstrated that engaging teachers in reflective prac-tice enhances their teaching effectiveness and provides opportunities for independent learning (Brookfield, 1986; Brookfield, 1995). In particular, Osterman (1990) argues that reflective practice "leads to greater self-awareness, to the development of new knowledge about professional practice, and to a broader understanding of the problems which confront practitioners" (p. 134), while also suggesting that this individual professional growth leads to organizational changes. In a review of the literature on change strategies in undergraduate STEM instruction, Henderson, Beach, and Finkelstein (2011) found that reflective practice was the only strat-egy that was commonly recognized as effective by the various types of change literature. Finally, reflective prac-tice has been demonstrated to be characteristic of effect-ive higher education instructors (Kane et al., 2004).

Although the literature on reflective teaching is largely focused on modeling processes of reflective practices and characterizing the extent to which instructors engage in these processes (Bubnys & Zavadskiene, 2017), several studies have also explored the quality of reflective practice (Hubball et al., 2005; Larrivee, 2008; Lane et al., 2014). One relatively recent study summarizes the literature and iden-tifies four levels (Larrivee, 2008): pre-reflection or non-reflection, surface reflection (reflection on strategies and methods without consideration of rationale for use), peda-gogical reflection (reflection includes knowledge and beliefs about best practices), and critical reflection (reflection on the morale and ethical implications of their teaching on students). Most of this work has been conducted with pre- and in-service teachers.

One aspect of reflective practice that has been investi-gated at the STEM postsecondary level is satisfaction with one's teaching. However, many of these studies only examine teaching satisfaction in the context of a single independent variable, such as an online versus brick-and-mortar classroom environment (Wasilik & Bolliger, 2009; Swartz, Cole, & Shelley, 2010). Gess-Newsome et al. (2003) differentiate between contextual dissatisfac-tion, a construct more closely tied to job satisfaction measures, and pedagogical dissatisfaction. The construct of pedagogical dissatisfaction or discontentment, explored more fully by Southerland and colleagues (2011a, b), is more relevant to the measures taken in our study. Pedagogical discontent refers to an instructor realizing that there is a misalignment between her/his teaching goals and his/her instructional practice: "peda-gogical discontentment is a teacher's affective response

to her evaluation of the effectiveness of her existing science teaching practices and goals" (Southerland et al., 2011a, p. 304). This dissonance can potentially result in the instructor being more willing to revise their practice and adopt new pedagogical approaches (Feldman, 2000). Studies conducted by Southerland and colleagues (2011b, 2012) identified different types of pedagogical discontent experienced by secondary science teachers: Ability to teach all students science, science content knowledge, balancing depth versus breadth of instruction, implementing inquiry instruction, and assessing science learning. In this study, we probed faculty's satisfaction with four types of teaching goals that relate to these observed areas of pedagogical discontent: achievement of learning goals, student learning, student engagement, and teaching strategy. Exploring faculty's pedagogical satisfaction provides us with a window into the support faculty would need to be fully engaged in reflective practice. We expected that limited reflections would lead to lack of pedagogical discontent.

Revisions
Stark (2000) and Hora and Hunter (2014) note that postsecondary course revisions often take the form of small modifications to previously used course materials, rather than wholescale overhauls or introductions of new teaching techniques. This type of thinking fits with the observation that changes made to instructional practices as part of a revision process are often filtered through a lens of how closely the new practice aligns with the previous one (Amundsen, Gryspeerdt, & Moxness, 1993). It is important to note that these discussions of instructional change and revision do not fully overlap with the common literature usage of "change" to represent higher-level changes to instructional practices (e.g., adoption of EBIPs, departmental and institutional transformation). Both lenses of instructional change are relevant to our study, but the former lens is particularly relevant to our research questions.

Research questions
The research questions addressed in this study are:

1. What pedagogical intentions did STEM faculty have for the teaching of a week of content?
   a. What types of learning goals do postsecondary STEM instructors have for their students?
   b. How do postsecondary STEM instructors plan to assess achievement of learning goals?
   c. What learning experiences do postsecondary instructors plan to use to help students achieve the learning goals?

2. To what extent were STEM faculty satisfied with their teaching of a week of content?
   a. To what extent are postsecondary instructors satisfied with their teaching?
   b. What types of revisions do postsecondary instructors plan to implement in the next execution of the course?
   c. What relationships exist between postsecondary instructors' level of satisfaction with their teaching and their intent for course revisions?

Methods
We employed a mixed methods design to answer our research questions. Mixed methods approaches have been described as "a pragmatic choice to address research problems through multiple methods with the goal of increasing the breadth, depth, and consistency of research findings" (Warfa, 2017). Mixed methods analysis does not inherently require the collection of both qualitative and quantitative data (Onwuegbuzie et al., 2007). In this study, we used a data-transformation mixed methods approach, where qualitative data was used both for qualitative analysis and quantitated for usage in quantitative analysis (Tashakkori and Teddlie, 1998). Critically, the usage of quantitative methods allowed us to go beyond the identification of codes within responses to individual interview questions, into testing for correlations between codes of interest relating to separate questions; these cross-code comparisons would have been unwieldy with qualitative findings alone.

Participants
This study characterizes a pool of 42 instructors drawn from a range of STEM disciplines at a Midwestern US public institution holding the Carnegie classification of "R1: Doctoral Universities—Highest Research Activity" (Table 1; Indiana University Center for Postsecondary Research, 2015). The enrollment is "high undergraduate," and undergraduate admissions are "more selective" with a low transfer-in rate (Indiana University Center for Postsecondary Research, 2015).

Three quarters of the studied instructors had voluntarily signed up for, but had not yet attended, an 8-week instructional workshop series addressing a specific instructional strategy—either Peer Instruction (PI; Mazur, 1997) or Just-in-Time Teaching (JiTT; Novak et al., 1999). The last quarter of these 42 instructors who had not enrolled in either of these workshops ("control") was recruited at the same time. Because the data was collected *prior* to workshop attendance, the treatment status does not necessarily indicate a difference in pedagogical training at the time of data collection, but instead signals future instructor intent. Workshop series addressing PI, JiTT,

Table 1 Attributes of instructors under study (*n* = 42)

Attribute		Percentage of instructors
Treatment status	"Control" (no workshop participation planned)	26
	Peer Instruction workshop attendance planned	43
	Just-in-Time Teaching workshop attendance planned	31
Class size	1 to 25	24
	26 to 50	19
	51 to 100	19
	101 through 150	12
	151 plus	26
Course level	Lower undergrad	62
	Upper undergrad	24
	Graduate	14
Course discipline	Biology	38
	Chemistry	24
	Math	10
	Physics	10
	Other (7 disciplines)	19
Experience (years)	0 to 6	21
	7 plus	79

or other topics relevant to STEM teaching have been offered at the study university since 2013. More than 100 faculty members (primarily tenured/tenure-track and a few lecturers) have attended one or more of the workshop series. A financial incentive of $500 was provided to participants who attended at least six of the eight workshop sessions.

Interview protocols and coding

Interviews were conducted both prior to and following a week of classes that served as the basis for questioning and discussion. The week of instruction was selected based on instructors' suggestions and research team availability; most took place between the first and last midterm examinations. Both interview protocols are found in the Supplementary Text (Interview protocols). Some interviews were conducted in person, and others were conducted via telephone, but all were audio-recorded.

The audio recordings of each paired interview (pre-teaching and post-teaching) were transcribed verbatim within a combined document, with vocal fillers removed. All authors contributed to the development of a coding scheme for the interviews, seeking information relating to the original research questions. The initial codebook consisted of 204 codes, including emergent codes as well as a small number of preconceived codes

(for example, the various levels of Bloom's taxonomy as applied to instructor learning goals). Descriptive coding was used to identify and group common practices and mindsets, magnitude coding was used to classify levels of instructor satisfaction, and structural coding was used to organize higher-order relationships between codes (Saldaña, 2013). Once an initial codebook was compiled, the interviews were coded using NVivo version 9 (QSR International, 2011). The unit of analysis during coding was the instructor's full response to an interviewer's question, with multiple codes applied to the same response when warranted. Five of the transcripts were coded by two authors, with the codes assigned by each author compared using pooled kappa to determine inter-rater reliability (IRR) (De Vries et al., 2008). The mean pooled kappa value for the five transcripts was 0.864, which indicates a high level of inter-rater reliability (Landis & Koch, 1977). Disagreements between the two coders within the five jointly coded transcripts were resolved to a consensus coding decision, with the discussion serving to further refine the operational definitions of the codes in question. The coding of the remaining 37 transcripts was split between the two authors who had coded the first five transcripts.

Following the initial coding process, we compressed the codebook through the merger of closely related codes and the elimination of minimally-informative codes (codes observed in less than 10% of relevant cases). We constructed a flowchart of initial to revised codes that allowed us to map mergers and changes from the initial coding scheme onto the revised coding scheme without any need for additional recoding. Through this process, we migrated from an initial codebook of 204 codes to a final codebook of 49 codes (Supplementary Tables 1-6). The frequencies reported within the results section reflect the number of participants whose responses were coded for the code in question at least once, as opposed to the total number of statements reflecting the code.

Cross-code comparisons

To investigate the significance of relationships between variables analyzed using contingency tables, we implemented Fisher's exact test using the command fisher.test within R (R Core Team, 2016). Fisher's exact test was chosen as a test that could be applied consistently across various sizes of contingency table, while allowing for low expected values within individual contingency table positions. When correcting for multiple comparisons, the Bonferroni method was utilized. In order to describe the magnitude of these correlations, we utilized the Goodman-Kruskal τ statistic (Goodman & Kruskal, 1954). τ provides an indication of

the reduction in error provided by one variable when used to predict a second variable. A value of 1 indicates that all error is eliminated (a variable enables perfect prediction of another variable) while a value of 0 indicates that no error is eliminated (a variable provides no information about another variable). It is an asymmetric measure, in that variable A may allow one to perfectly predict variable B without variable B providing an equivalent level of information about variable A. τ values were calculated using the GoodmanKruskal package (Pearson, 2016) within R (R Core Team, 2016).

Results

In this study, we were interested in capturing elements of instructors' thinking when planning a week of teaching and identifying the relationships between their level of satisfaction with the week of teaching and intent for course revisions. A description of each participant quoted or mentioned in this section can be found in the Supplementary Table 7; all names used below are pseudonyms.

Identifying learning goals

When instructors were asked what their learning goals were for the upcoming week of instruction, almost all of the instructors (98%) listed the topics they were planning to cover (Fig. 1a). A subset of instructors also brought up skills development (14%) and/or real-world connections (12%) as learning goals.

Most instructors (88% of participants) also presented learning goals that could be classified using the revised Bloom's taxonomy (Anderson & Krathwohl, 2001; Fig. 1a). For this study, *evaluation*, *synthesis*, and *analysis* learning goals were considered higher-level thinking processes and were labeled higher-level goals, while *application*, *knowledge*, and *comprehension* learning goals were considered lower-level thinking processes and labeled lower-level goals. As shown in Fig. 1a, instructors mostly listed lower-level goals. Indeed, all instructors who mentioned learning goals that could be classified with Bloom's taxonomy included at least one lower-level goal. For example, Jim, a lower-level undergraduate Biology instructor, stated:

> …they have to know the differences between the prokaryotic cell and the eukaryotic cell. So [students] say some [cell structures], and oh this belongs to prokaryotic cells, and these belong to eukaryotic cells, and then call it that. When we talk about the functions of the cells, they will know that this function is which part of the cells, and talk about the organelles or structures of the cell that are responsible for each important function of cells. And the following lecture I will

> talk about work in cells, meaning about energy, about how a cell has this energy to work…

By asking students to identify cell structures in order to categorize the cells, to remember the differences between prokaryotic and eukaryotic cells and to grasp the importance of energy in cell functioning, Jim has set knowledge and comprehension learning goals for his students.

More cognitively demanding learning goals that required students to analyze, synthesize, or evaluate were mentioned by only 17% of the study's participants. For example, one physics instructor, Michael, described the overarching goals of his course as:

> wanting students to stop working with algorithmic problem solving and prescription-based problem solving and move to solving complex problems based on their understanding of the system. I always want students not to solve problems step by step, but instead solve problems by understanding the system and building up an analytical description of the system.

Here, Michael is asking his students to critically analyze the physics problem, apply the appropriate physics concepts, and then develop a method to solve the problem using the concepts. By making this method of problem solving his overarching course objective, Michael demonstrates that he wants his students to engage in higher-order thinking processes.

Assessing learning goal achievement

Instructors were then asked how they planned to assess their students' achievement of their learning goals for that week of teaching. Figure 1b shows the assessment strategies mentioned by at least 20% of the study participants.

Instructors mostly used summative assessment methods to evaluate student learning, including exams (81% of instructors) and quizzes (48%). A smaller subset of instructors used formative assessment strategies such as clicker questions (38%), student questions and comments (38%), student responses to instructor questions (31%), and in-class assessment activities (24%). Instructors relied extensively on homework as well (69%); however, it was not always clear if homework were utilized formatively or summatively. On an individual instructor level, these frequencies generally translate into the mention of multiple summative assessment modalities paired with a lesser number of formative assessments. Furthermore, instructors often indicated that summative assessments were the default or expected option. For instance, Bob, who instructs upper-level undergraduate earth and atmospheric sciences, explained his assessment practices with the following:
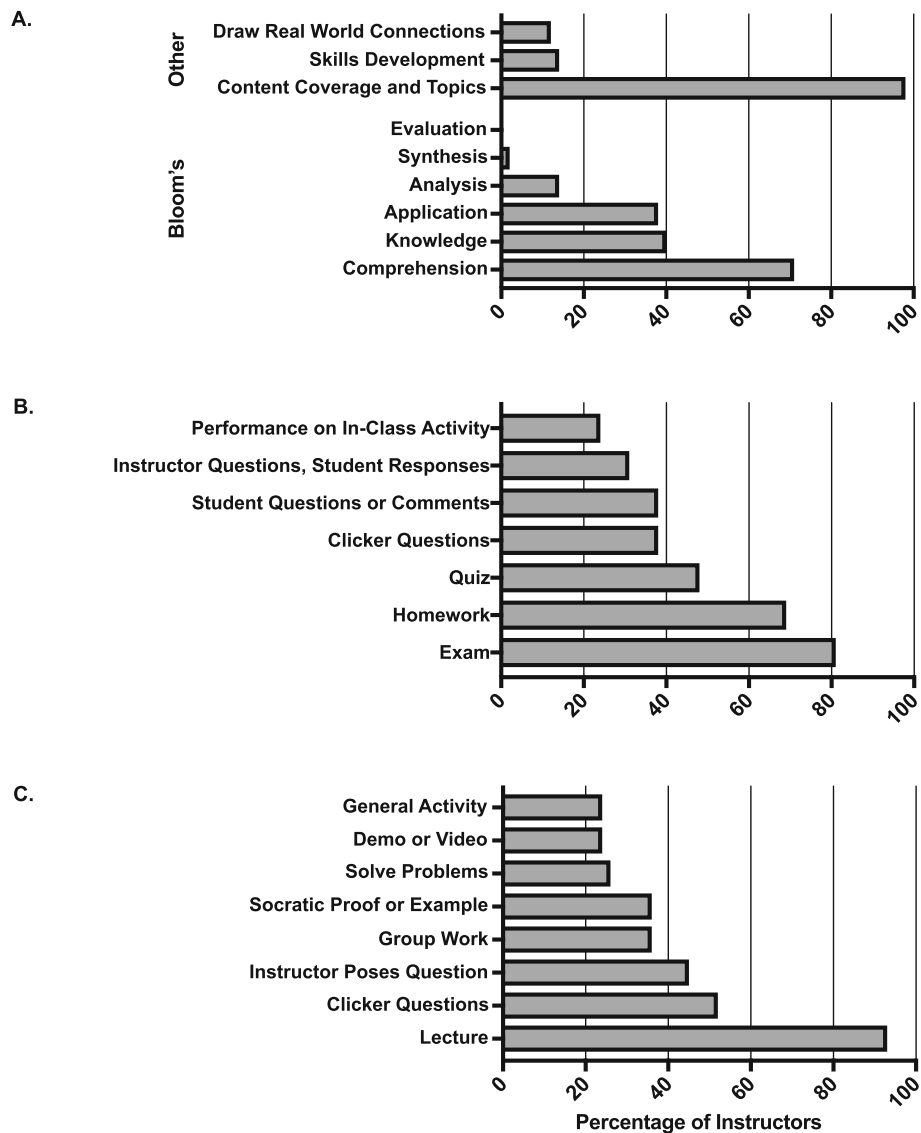
Fig. 1 Instructor descriptions of the planning process. Codes that emerged from participants' answers to questions 3–4 in the pre-interview. **a** Learning goals. **b** Assessment strategies. **c** Engagement strategies. The frequencies reflect the number of participants whose responses were coded for the code in question at least once, as opposed to the total number of statements reflecting the code

I guess there is kind of formal and informal. Informally, we will have some activities in class so I can see how they discuss these concepts, see how they understand, so I will get an initial assessment. Then more formally, we have quizzes and lab exercises, and, of course, exams and such.

Planned learning experiences

During the interview prior to the observed week of teaching, instructors were asked to explain how they plan to engage students in their class that week. Figure 1c displays the engagement strategies mentioned by at least 20% of the participants.

Almost all instructors (93% of participants) reported using traditional lecturing to engage their students. Of the 39 instructors who reported using lecture as an engagement method, 97% also reported using some form of questioning—either verbal or through clicker questions—interspersed in their lectures. The three instructors who did not mention lecture as an engagement strategy used a number of other approaches, mentioning three, four, and six non-lecture engagement strategies respectively. Overall, most instructors reported a varied selection of engagement strategies, with 90% of participants mentioning at least three engagement strategies, and 60% describing at least four separate strategies.

Interestingly, we found that strategies which are aimed at both engaging and assessing students were rarely mentioned as formative assessment tools. For example, although 52% of participants mentioned using clicker questions in class to engage students (Fig. 1c), only 38% of that subset identified clicker questions as an assessment tool. Only one instructor mentioned clicker questions solely as an assessment strategy. Angela, a lower-level undergraduate biology instructor who mentioned using clicker questions as a form of student engagement, said the following regarding assessment:

I don't really use clicker questions to assess their learning. [Students] use clicker questions to assess their learning, and I use my lecturing. I assess their learning on exams. I don't really care if they get the clicker questions right or not, as long as they are participating.

Angela's thoughts about using clicker questions for engagement but not assessment were mirrored by eight of her colleagues, representing 21% of the participants in total. Another engagement strategy that has the potential to be used as a formative assessment method is just-in-time teaching (JiTT). JiTT was mentioned by six participants as an engagement strategy, but by only two participants as an assessment strategy.

Satisfaction with teaching, goal achievement, and student engagement

During the interview following the week-long teaching period, instructors were asked to express the extent to which they were satisfied with goal achievement, student learning, their teaching, and student engagement. Results are presented in Fig. 2a.

Satisfaction with goal achievement

Overall, participants were satisfied with the attainment of their goals for their classes, with over half of the participants (60%) being somewhat satisfied and 29% reporting high satisfaction. Instructors based their assessment on several different methods: content coverage (mentioned by 52% of participants), formative and summative assessment results (33%), personal feelings (31%), and student engagement and attention levels (14%). For example, Meredith, an instructor of an upper-level undergraduate biology course who was somewhat satisfied with her goal achievement, cited content coverage as the underlying reason for her satisfaction level:

I would say I met 70% of my goals. The presentation of the background material took longer than I anticipated and I'm usually overestimating how much material I can cover.

All four instructors who were dissatisfied with their goal achievement mentioned content coverage as their underlying reasoning. For example, Andy, who teaches a lower-level mathematics course, stated:

I think I struggled a little bit to get the material...it was more rushed than I wanted it to be, so I feel I struggled a bit to meet those goals. So I am not sure I was completely successful with them.

At the time of the interview, one instructor, Holly, was unable to provide an assessment of her satisfaction because she had not yet assessed her students:

I won't really [know] until the kids are assessed... I'm not sure...Once I ask them the question and I see if they got it, then I'll have a better idea of how well things went.

Interestingly, this post-teaching comment aligns with her pre-teaching interview in which she indicated only using exams to assess her students' learning. Even though she was implementing clicker questions, she saw them as a tool to engage students rather than assess them.

Four instructors stated that they were highly satisfied or somewhat satisfied with their goal achievement while simultaneously indicating a degree of uncertainty about their satisfaction level. Like Holly, two of these instructors were waiting until they had results from summative assessments to make a definitive statement. For example, Toby, an instructor of a lower-level biology course, explained:

I would say, it's difficult to answer that question until I see how they do on the exam and homework and such...But as far as accomplishing what I set out to do, I guess I'm a little bit behind schedule, but other than that, yeah, it was fine.

The other two instructors, who were uncertain about their satisfaction level, were having difficulty recalling the learning goals they had mentioned in the pre-teaching interview.

Student learning

Following the week of instruction, instructors were asked whether they thought students learned what they intended them to learn. Out of the 40 instructors who were asked this question, 30% reported not knowing yet at the time of the interview, 42% reported measures of student learning, while the remaining 28% shared some measure of student learning while simultaneously referencing assessment results they did not yet have available. None of the participants reported that their students did
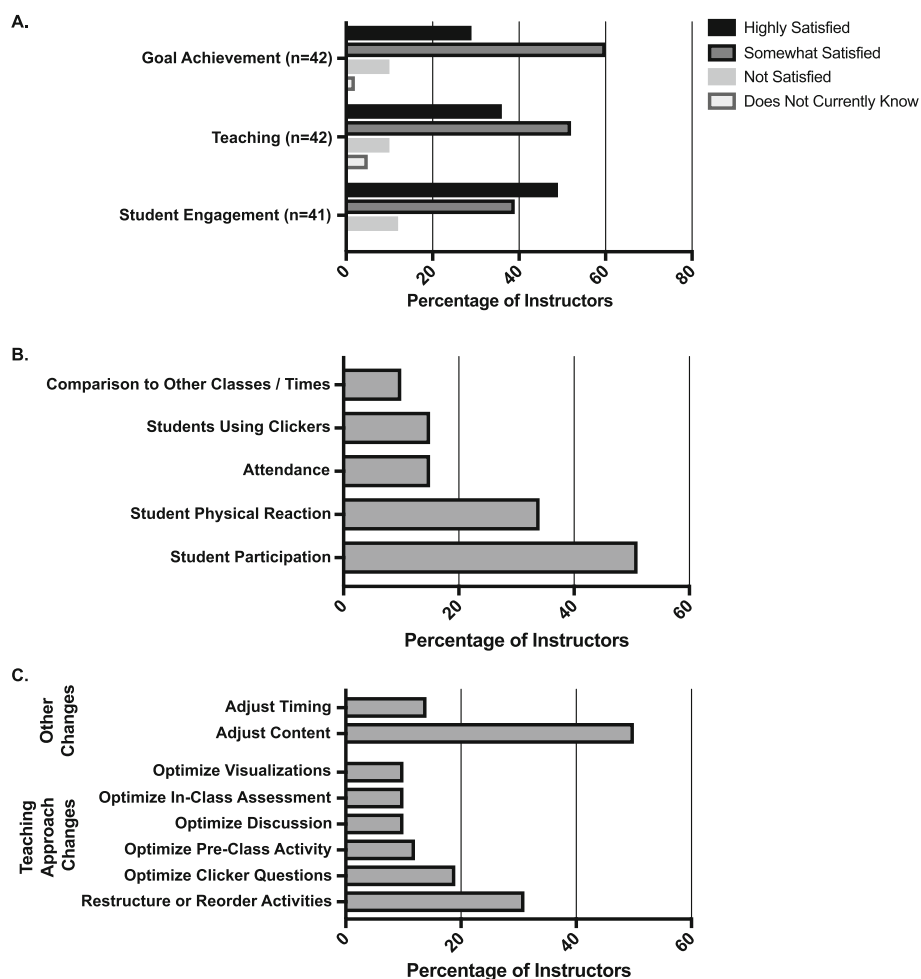
Fig. 2 Instructor reflections. Percentage of instructors reporting in their interview on their **a** level of satisfaction with teaching, learning goal achievement, and student engagement, **b** measurement methods that they used to gauge student engagement within their classrooms, and **c** changes they plan to make when teaching the observed week of classes in the future (only including the instructors that indicated at least one planned change in their response)

not learn overall, even if some went on to mention particular assessment strategies where students did not perform particularly well.

Among the 23 instructors who mentioned that they were waiting on at least some measures of student learning, the majority (61%) referenced exams, showing a clear emphasis on summative assessment techniques (Table 2). Indeed, nine of the 12 instructors who reported not yet having any available information about student learning simply stated that they were waiting upon the exam for information about student learning, with no other potential measures mentioned. For example, Dwight, an instructor of a lower-level biology course, said,

I really won't know until I give them the exam and at the exam time, I really don't go back and see who was actually in attendance and did well on the exam…

Meanwhile, the 28 instructors who mentioned that they had at least some measure of student learning available primarily relied on formative assessment methods, such as clicker questions, responses to instructor questions, and in-class activities, to justify their response.

Teaching satisfaction

With respect to teaching, instructors were largely satisfied, with 52% of the participants being somewhat satisfied and 36% reporting high satisfaction. Participants mentioned personal feelings (43%), student engagement (36%), and assessment results (12%) as reasoning for their level of teaching satisfaction. For example, Jan, an instructor of a lower-level undergraduate chemistry course, was highly satisfied with her teaching and based her evaluation on personal feelings about the class:

Table 2 Methods used to determine student learning

Method used, or to be used, to determine student learning	Instructor knows if students learned	Instructor does not know yet	Total
Exam	1	14	15
Clicker questions	8	3	11
Homework	3	7	10
Responses to instructor's questions	10	0	10
In-class activities/assessment	8	1	9
Quiz	5	2	7
Student questions or comments	5	1	6
JiTT	2	0	2
Muddiest point	2	0	2

Following a week of instruction, instructors were asked whether or not their students had learned the content they had taught. For those who responded that their students had learned the content, the "Instructor knows if students learned" column indicates what type of evidence the instructor cited (with instructors able to cite more than one method). The "Instructor does not know yet" column breaks down the methods cited by instructors who responded to this question by indicating that they were still waiting on some measure of student learning not yet available

I was very satisfied... I mean, it can always be better, but with the amount of time I have, I think I use most of the tools that we have, like using the clicker, using the PowerPoint, using models to give students various ways to learn the same thing... I'm sure if someone else sees it, they might say this could be better, but I feel I've done my best.

Two instructors reported not currently knowing the extent to which they are satisfied with their teaching; one of the two instructors, Andy, mentioned personal feelings, student engagement levels, and student difficulties as reasons for his uncertainty:

I feel very equivocal about it. I believe the message is sound, but I am not sure that I am executing [the lectures] as efficiently as I should be. Keeping people on task...there is always a trade-off between efficiency and understanding, but I think I am getting burned on groups getting off task... the classes bifurcate and some groups are getting the material and understanding it and some groups are struggling to get the material in a reasonable way...

A total of four instructors were dissatisfied with their teaching. These instructors also cited their personal feelings and classroom engagement levels as reasons for their dissatisfaction.

Satisfaction with student engagement

Compared to teaching and goal achievement, instructors were generally more satisfied with their students' engagement. As shown in Fig. 2a, 49% of the instructors were highly satisfied with engagement in their classrooms, while 39% were somewhat satisfied with engagement. Most instructors (80%) mentioned students' behaviors such as participation levels (51% of participants), students' physical reactions (34%), and attendance (15%) to support their assessment of student engagement (Fig. 2b). For instance, Clark, an experienced physics instructor who was highly satisfied with student engagement, relied on students' physical reactions:

Whenever I teach one of these big introductory courses, the students are quite engaged, people aren't falling asleep and reading the [school newspaper] and so far they seem to be paying attention to me... You can look at the eyes of 150 students in a broad sweep, and if you just said something that doesn't resonate or sink in, you get this kind of average glazed over look of the whole class...The students are engaged enough that I can tell from the way they are looking at me, just the eye contact that I'm making in this big lecture format, whether they are getting it or not, the people seem to be quite engaged.

For instructors who used the participation of their students, some based their judgement on a couple of interested students, whereas others remarked on broad student participation.

Out of the 41 instructors who were asked this question, five reported that they were dissatisfied with student engagement. Two of these instructors did not provide a descriptive explanation for why they felt their students were not engaged. The remaining three all cited low levels of student attention and/or participation as a factor in their discontentment. Stanley, who teaches lower-level biology courses, was dissatisfied with student engagement and reported that his class was less engaged than normal due to his teaching methods. He said:

I didn't have very much interactive stuff. I had more of it last week than I did this week. I think it may be partly the material. I don't really know. I just had the sense that they weren't as into it, I guess...

Revisions planned for future iterations of course

Participants were asked what they would do differently if they were to teach the week of content again in the future. The changes that instructors indicated are presented in Fig. 2c.

Only four of the instructors expressed no desire to change future iterations of the focal week of content. Over half of the participants (51%) suggested that they would adjust their content coverage by either leaving content out, including more content, or talking about material at a different time point during the semester. Meanwhile, 73% of instructors stated that they would like to change their approaches to teaching, with the most frequently suggested change of approach (32%) being a restructuring of in-class activities. Yolanda, an instructor of a graduate-level entomology course, stated:

> I think what I'd like to try is have them do a little bit more outside work - not necessarily flip the class, but do a little bit more blended-type learning so we do a little bit more outside the class and maybe do, instead of quizzes and study sets, which are kind of after the fact, I'd like to try more of the JiTT approach where they essentially read the material before class and I am able to assess where they stand with the information. So if they understand something pretty well, then there is no reason for me to go in, but if they are having trouble with it then I can spend time with it. So of course it's a much more dynamic way of teaching, and I don't know if it's going to work because we cover so much material, but maybe specific topics per se, maybe I can do it that way and that's what I was thinking of.

Some instructors also mentioned improving the quality and increasing the quantity of instructional activities. In particular, 20% of the participants mentioned optimizing clicker questions, 12% reported wanting to optimize their pre-class activity, and 10% reported a desire to optimize their in-class discussions. For instance, Jan said:

> I didn't have any clicker questions because that one that I had prepared, came at the very end; I had no time…So maybe I would have some clicker questions based on some basic material, other than the advanced question that I didn't get to before. A little earlier, maybe [I will include] easier low-level questions. And that way I would know if the students understood the basics before getting to the advanced one. So just tailoring the clicker questions a bit or something like that.

When instructors offered reasoning for their course changes, they most often mentioned optimizing student learning (36%) or increasing student engagement (31%).

Relationship between instructor satisfaction and course revisions

Instructor satisfaction and planning of course revisions

We then examined the degree of correlation for the relationships between instructor satisfaction and course revisions codes, and failed to find any significant correlations. For example, Supplementary Table 8 illustrates that the frequency of change to approaches to teaching has no significant connection to the level of instructor satisfaction with goal achievement ($p = 0.47$, two-tailed Fisher's exact test). We then expanded beyond point comparisons and looked at the connections between the three types of satisfaction and a set of four codes relating to course revisions (Table 3). None of the 12 comparisons rose to the level of statistical significance with a two-tailed Fisher's exact test, signaling no more than a minimal relationship between the level of instructor satisfaction and instructional revisions within our population.

This was further confirmed by analyzing instructors who expressed full satisfaction in all three aspects (goal achievement, teaching, and student engagement). Three out of 42 instructors met this criterion. One of these three, Dwight expressed no desire to make revisions. However, the other two each identified multiple areas where they wanted to change their instructional approaches. Darryl, a lower-level undergraduate mathematics instructor, mentioned alterations to learning goals as well as plans to adjust content and teaching approaches. Meanwhile, Tony, a graduate chemistry instructor, mentioned alterations to timing of instruction and adjustments to teaching approaches. These cases indicate that instructors who express broad-based satisfaction with

Table 3 Significance of correlations between course revisions codes and other variables

		Satisfaction with teaching	Satisfaction with engagement	Satisfaction with goal achievement	Experience level
Types of course revisions	Change approaches to teaching	ns	ns	ns	ns
	Adjust classroom timing	ns	ns	ns	ns
	Adjust course content	ns	ns	ns	$p < 0.0013$
	Change course goals	ns	ns	ns	$p < 0.0004$

ns not significant

various aspects of their instruction still have a desire and concrete plans to adjust their practices.

Drivers of course revisions

Since satisfaction levels were poor predictors of planned course revisions, we looked for codes or attributes showing higher levels of correlation. Nearly all of the instructor attributes we investigated (e.g., treatment status, course level, course discipline, class size) were less statistically associated with course revisions than instructor satisfaction, with the exception of instructor experience. We categorized instructors as inexperienced if they had six or fewer years of teaching experience and as experienced if they had seven or more years of teaching experience. This division was based on a number of factors: pre-tenured faculty are more likely to receive feedback from senior colleagues as part of the review process which may also trigger further modification of their practices (peer review of teaching is mostly absent post-tenure at this institution), and by the time they are promoted and tenured, they have taught most of their courses at least twice and are likely to feel more settled. We found that inexperienced instructors show a greater propensity to express a desire to change the content that they will teach in future courses (p = 0.0013, two-tailed Fisher's exact test with Bonferroni multiple comparisons correction of α = 0.0125; τ values are 0.273/0.273; Supplementary Table 9). We further examined the connection between teaching experience and the set of four course revisions codes utilized above (Table 3). Two of the four comparisons were statistically significant, and in both cases, the more experienced instructor group was less likely to express a desire to change course content or course goals. The average τ value for such ties is 0.151, showing that the correlation between course revisions and experience level is greater in magnitude than the correlation between course revisions and satisfaction level. Taken together, it appears that instructor experience may play an important role within the instructional change processes of instructors.

Discussion and implications

Instructional planning is instructor-centric

Our findings point to an instructor-centric planning mindset with limited considerations of student learning. Many of our results are in alignment with results of prior studies. With regards to learning goals, we observed that most of our participants identified lower-level learning goals and were quite content-centric in their descriptions (Fig. 1a). This fits with the findings of Stark (2000), who observed that by far the most frequent first step taken by instructors planning courses across a range of disciplines is the selection of content. Meanwhile, the distribution of Bloom's levels for participants' learning goals was quite similar to the distribution observed in an analysis of dozens of college biology syllabi

(Momsen et al., 2010). In terms of assessment practices, the majority of our participants utilized summative assessment techniques, while fewer utilized formative assessment techniques (Fig. 1b). This differential usage pattern dovetails with the findings of national surveys of instructors within the field of geoscience (Macdonald et al., 2005). With regards to classroom practice, we found that although two-thirds of the participants integrated student-centered activities into their teaching (i.e., clicker and group work; Fig. 1c), it was common for them to report lecture as a primary mechanism when we asked about methods of *engaging* their students (as opposed to asking what the instructor would be doing). This also fits with a number of recent studies that have characterized the heavy utilization of lecture-based instruction within higher education STEM classrooms (Apkarian & Kirin, 2017; Stains et al., 2018; Teasdale et al., 2017).

Instructors see formative assessment tools as engagement tools rather than assessment tools

The results detailed in Fig. 1b, c and Table 2 come together to illustrate the complexities of characterizing instructors' understandings and mindsets regarding assessment. In post-teaching interviews, clicker questions were cited with the second highest frequency as a method of determining student learning. The 11 instructors who mentioned clicker questions as a way of determining student learning in the post-teaching interviews all listed clickers as both an engagement technique and an assessment method in the pre-teaching interview. However, the percentage of instructors who reported using clicker questions as an engagement technique is notably higher than the percentage of instructors who reported using clicker questions as an assessment technique in pre-teaching interviews. This raises questions about how exactly instructors conceptualize the balance between the assessment and engagement functions of various formative assessment methods. It also speaks to the need for professional development to focus on helping STEM faculty develop an awareness and understanding of the role of clickers and potentially other formative assessments in improving instructional effectiveness.

STEM faculty minimally rely on indicators of student learning when evaluating their level of satisfaction with their teaching

As described in the "Results" section regarding learning goal satisfaction and teaching satisfaction, a plurality of instructors (43%) mentioned personal feelings as a major determinant for how satisfied they were regarding their teaching. Only 12% of instructors brought up any form of assessment data when describing their levels of teaching satisfaction. In the case of satisfaction with goal

achievement, the fraction of instructors citing formative and/or summative assessment results (33%) and personal feelings (31%) are nearly identical. This means that personal feelings are considered by many instructors when they are thinking about their teaching.

Interestingly, instructors in this study were not clear about student learning at the end of the week of instruction. Indeed, about a third of the faculty did not know whether their students had learned the targeted learning goals and another third had some evidence but were also waiting on the results of summative assessments. None of them said that students did not learn, including those instructors that indicated that their assessment results were not at the desired level of performance.

The reliance on personal feelings and the lack of attention to student learning indicates that most STEM faculty in this sample had a level of reflection that would be described as non-reflective by studies investigating quality of reflective practice (Lee, 2005; Larrivee, 2008; Lane et al., 2014). This finding illustrates the need for professional development programs to explicitly teach STEM faculty about effective and meaningful reflective practice (Russell, 2006) and for institutions of higher education to value that practice. Importantly, reflective practice should not be taught and evaluated as a mere set of processes but should be embedded in a comprehensive pedagogical training program and institutional approaches on teaching that focus on evidence-based practices for learning (Edwards & Thomas, 2010).

Pedagogical discontent was not a driver for course revisions

Prior work has argued that pedagogical discontent is necessary (but insufficient) for the implementation of new instructional strategies (Andrews & Lemons, 2015). However, we found that the three types of instructor satisfaction we assayed were surprisingly decoupled from instructors' responses regarding desire or plans to change instructional practices in the future. Instead, we found that instructor change codes were more tightly coupled with the level of instructor experience. This lack of connection between discontent and curricular revisions may reflect instructors' limited engagement in reflective practice since discontent may only be achieved if one is engaged in in-depth reflections.

When considering the connections between instructor change and instructor satisfaction, it is important to bear in mind the potential for not-comparable definitions of change. In the case of instructional change, there is a difference between defining change as small adjustments to teaching a course, the implementation of a single target teaching strategy, and a complete course overhaul. In our study, in contrast to other studies, we focused on small adjustments that faculty make to their course.

Pedagogical discontent may thus be critical for faculty to engage in practices that depart significantly from their current practices (i.e., conceptual change; Feldman, 2000) but not for course revisions.

Similarly, instructor satisfaction has the potential to be interpreted in different fashions. Contrary to studies on postsecondary instructors in which pedagogical discontent was treated as a single construct, we leveraged the K-12 literature (Southerland et al., 2011a, b) and subdivided it into related, but distinct constructs. Our data indicate that instructors often respond to inquiries about various aspects of their satisfaction in different ways, suggesting that boiling down satisfaction to a single value risks oversimplifying the complexities of instructor mindsets.

We observed multiple examples of highly satisfied instructors expressing a desire to implement new teaching strategies in future iterations of their classes. This unexpected result leads us to advocate the further study of potential relationships between dissatisfaction and instructional change, especially in light of the differential approaches used by our study compared to others that have explored the connections between these constructs.

Professional development need to align with instructors' mindset

Results of this study further highlight the critical need to investigate STEM faculty's instructional decision and value system and to incorporate these findings into instructional reform efforts. For example, our results on instructional planning indicate that STEM faculty do not necessarily focus on student learning when reflecting on and evaluating their teaching practices. Yet, much of the rhetoric of the discipline-based education research community and the associated professional development community has been to promote active learning and evidence-based practices to faculty as a mean to address the weak learning performance of their students (Freeman et al., 2014). This argument would fail for many faculty in our sample since they minimally considered student learning but rather focused on their personal feelings. We need to identify ways to scaffold faculty's learning and implementation of best instructional practices from where they are—focused on personal feelings rather than student learning—rather than providing prescribed solutions that do not fit their mindsets. Moreover, timing seems to be critical. Since, less experienced instructors seem to be more willing to revise their pedagogical approach than seasoned instructors, future studies should investigate when the most influential time is in a new faculty's teaching career to provide them with pedagogical professional development.

Limitations

This study was undertaken at a single doctorate-granting university in the USA, which places constraints on its generalizability to the entire higher education instructional pool. Replications performed in similar and dissimilar settings may provide the ability to extrapolate findings into broader contexts. In addition, the "control" and "treatment" groups were of different sizes and showed clear differences in teaching satisfaction. It is difficult to know how much of a total pool of instructors would fit into a "control" versus "treatment" mentality overall, so it is possible that we are oversampling from a subpopulation that is more open to instructional change than the global average. However, the relative paucity of meaningful differences found between "control" and "treatment" instructors on a broad spectrum of analyzed codes may moderate this concern.

We did not collect data from participating instructors regarding their background pedagogical knowledge, training, or experiences. It is possible that the range of responses that any individual instructor could conceivably give might have been limited by a relative lack of exposure to various instructional practices. For instance, it is harder to imagine changing one's instructional practices if one is not aware that alternative instructional practices exist. This represents a "hidden" variable that might have been able to contextualize some components of our results.

Conclusions

In this study, we explored the pedagogical intentions of STEM instructors employed by a single doctorate-granting institution in the USA. Our results highlight that STEM faculty's instructional intentions are heavily focused on content coverage and personal feelings and only minimally involve students' learning outcomes. We also found that STEM faculty have limited knowledge of the role of assessment, especially formative assessment, in informing their own practices. When examining the relationships between pedagogical discontent and revisions of instructional practice, we discovered that instructor satisfaction levels for any segment of pedagogical practice were a poor predictor of intent for future instructional revisions. Instead, revision intent was most accurately predicted by the instructor experience level.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40594-020-00206-7.

Additional file 1. Supplementary Text Interview protocols.

Supplementary Table 1:. Codes related to learning goals.

Supplementary Table 2:. Codes related to assessment.

Supplementary Table 3:. Codes related to engagement strategies .

Supplementary Table 4:. Codes related to satisfaction.

Supplementary Table 5:. Codes for measurement of instructor satisfaction.

Supplementary Table 6:. Codes for course revisions.

Supplementary Table 7:. Characteristics of participants who were quoted or mentioned in the results section.

Supplementary Table 8:. Cross-comparison of goal achievement satisfaction with desire to change teaching.

Supplementary Table 9:. Cross-comparison of content adjustment with instructor experience level.

Abbreviations
EBIP: Evidence-based instructional practice; JiTT: Just-in-time teaching; PI: Peer instruction; STEM: Science, Technology, Engineering, Mathematics

Authors' contributions
KM and RE coded and analyzed the interview transcripts and contributed to the writing of the manuscript. MS designed the study, contributed to data collection and analysis, as well as the writing of the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
This work was approved by the Institutional Review Board at the University of Nebraska-Lincoln (IRB Number: 20120712790 EX). All participants were consented prior to taking part in the study.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Campus Learning Data and Technology, University of Minnesota Rochester, Rochester, MN, USA. [2]William Jennings Bryan High School, Omaha, NE, USA. [3]Department of Chemistry, University of Virginia, Charlottesville, VA 22904-4319, USA.

References
American Association for Higher Education. (2000). *Targeting curricular change: reform in undergraduate education in science, math, engineering, and technology*. Washington, D.C.: American Association for Higher Education.

Amundsen, C., Gryspeerdt, D., & Moxness, K. (1993). Practice-centered inquiry: Developing more effective teaching. *Review of Higher Education, 16*(3), 329–353.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: a revision of bloom's taxonomy of educational objectives*. New York: Longman.

Andrews, T. C., & Lemons, P. P. (2015). It's personal: Biology instructors prioritize personal evidence over empirical evidence in teaching decisions. *CBE - Life Sciences Education, 14*(1), 1–18.

Apkarian, N., & Kirin, D. (2017). Progress through calculus: census survey technical report. Retrieved from https://www.maa.org/sites/default/files/PtC%2 0Technical%20Report_Final.pdf. Accessed 9 May, 2018.

Association of American Universities. (2017). Undergraduate STEM Education Initiative. https://www.aau.edu/education-service/undergraduate-education/ undergraduate-stem-education-initiative. Accessed 24 April, 2017.

Bloom, B. S., Engelhart, M. D., Furst, F. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain*. New York: David McKay.

Borrego, M., Froyd, J. E., & Hall, T. S. (2010). Diffusion of engineering education innovations: A survey of awareness and adoption rates in U.S. engineering departments. *Journal of Engineering Education, 99*(3), 185–207.

Borrego, M., & Henderson, C. (2014). Theoretical perspectives on change in STEM higher education and their implications for engineering education research and practice. *Journal of Engineering Education, 103*(1), 45–76.

Bretz, S. L. (2012). Navigating the landscape of assessment. *Journal of Chemical Education, 89*(6), 689–691.

Brookfield, S. D. (1986). *Understanding and facilitating adult learning: a comprehensive analysis of principles and effective practices*. London: McGraw-Hill Education.

Brookfield, S. D. (1995). *Becoming a critically reflective teacher*. San Francisco: Jossey-Bass.

Bubnys, R., & Zavadskiene, L. (2017). Exploring the concept of reflective practice in the context of student-centered teacher education. *Proceedings of the International Scientific Conference, 1*, 91-101.

Clark, C., & Lampert, M. (1986). The Study of Teacher Thinking: Implications for Teacher Education. *Journal of Teacher Education, 37*(5), 27–31.

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 255–296). New York, NY: MacMillan.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and teacher education, 18*(8), 947–967.

Davila, K., & Talanquer, V. (2010). Classifying end-of-chapter questions and problems for selected general chemistry textbooks used in the United States. *Journal of Chemical Education, 87*(1), 97–101.

De Vries, H., Elliott, M. N., Kanouse, D. E., & Teleki, S. S. (2008). Using pooled kappa to summarize interrater agreement across many items. *Field Methods, 20*(3), 272–282.

Eagan, K. (2016). Becoming More Student-Centered? An examination of faculty teaching practices across STEM and non-STEM Disciplines between 2004 and 2014: Alfred P. Sloan Foundation.

Edwards, G., & Thomas, G. (2010). Can reflective practice be taught? *Educational Studies, 36*(4), 403–414.

Emenike, M., Raker, J. R., & Holme, T. (2013). Validating chemistry faculty members' self-reported familiarity with assessment terminology. *Journal of Chemical Education, 90*(9), 1130–1136.

Feldman, A. (2000). Decision making in the practical domain: a model of practical conceptual change. *Science & Education, 84*(5), 606–623.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, 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.

Geis, G.L. (1996). Planning and developing effective courses. In R.J. Menges, M. Weimer, & Associates, Teaching on solid ground: using scholarship to improve practice. San Francisco: Jossey-Bass.

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). New York: Routledge.

Gess-Newsome, J., Southerland, S. A., Johnston, A., & Woodbury, S. (2003). Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching. *American Educational Research Journal, 40*(3), 731–767.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49*(268), 732–764.

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

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*, 020102.

Henderson, C., & Dancy, M. H. (2009). Impact of physics education research on the teaching of introductory quantitative physics in the United States. *Physical Review Special Topics-Physics Education Research, 5*(2), 020107.

Hora, M. T., & Hunter, A.-B. (2014). Exploring the dynamics of organizational learning: identifying the decision chains science and math faculty use to plan and teach undergraduate courses. *International Journal of STEM Education, 1*, 8.

Hubball, H., Collins, J., & Pratt, D. (2005). Enhancing reflective teaching practices: implications for faculty development programs. *The Canadian Journal of Higher Education, 35*(3), 57–81.

Indiana University Center for Postsecondary Research. (2015). *The Carnegie classification of institutions of higher education*. Bloomington, IN: Indiana University Center for Postsecondary Research.

Kane, R., Sandretto, S., & Heath, C. (2004). An investigation into excellent tertiary teaching: Emphasising reflective practice. *Higher Education, 47*(3), 283–310.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lane, R., McMaster, H., Adnum, J., & Cavanagh, M. (2014). Quality reflective practice in teacher education: a journey towards shared understanding. *Reflective Practice, 15*(4), 481–494.

Larrivee, B. (2008). Development of a tool to assess teachers' level of reflective practice. *Reflective Practice, 9*(3), 341–360.

Lee, H.-J. (2005). Understanding and assessing preservice teachers' reflective thinking. *Teaching and Teacher Education, 21*(6), 699–715.

Lund, T. J., & Stains, M. (2015). The importance of context: an exploration of factors influencing the adoption of student-centered teaching among chemistry, biology, and physics faculty. *International Journal of STEM Education, 2*(1), 13.

Macdonald, R. H., Manduca, C. A., Mogk, D. W., & Tewksbury, B. J. (2005). Teaching methods in undergraduate geoscience courses: Results of the 2004 on the cutting edge survey of U.S. faculty. *Journal of Geoscience Education, 53*(3), 237–252.

Mazur, E. (1997). *Peer Instruction: A User's Manual*. Upper Saddle River, NJ: Prentice Hall.

Menges, R. J., & Austin, A. E. (2001). Teaching in higher education. In V. Richardson (Ed.), *Handbook of research on teaching*. Washington, DC: American Educational Research Association.

Mervis, J. (2014). Studies suggest two-way street for science majors. *Science, 343*, 125–126.

Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the Facts? Introductory Undergraduate Biology Courses Focus on Low-Level Cognitive Skills. *CBE - Life Sciences Education, 9*(4), 435–440.

National Research Council. (1999). *Transforming undergraduate education in science, mathematics, engineering, and technology*. Washington, DC: National Academy Press.

National Research Council. (2003a). *Bio 2010: Transforming undergraduate education for future research biologists*. Washington, DC: The National Academies Press.

National Research Council. (2003b). *Evaluating and improving undergraduate teaching in science, technology, engineering, and mathematics*. Washington, D. C.: National Academy Press.

National Research Council. (2011). *Promising practices in undergraduate science, Technology, Engineering, and Mathematics Education: Summary of Two Workshops*. Washington, DC: The National Academies Press.

National Research Council. (2012). *Discipline-based education research: understanding and improving learning in undergraduate science and engineering*. Washington, D.C.: The National Academies Press.

National Science Foundation. (2013). Widening implementation & demonstration of evidence-based reforms (WIDER). Program Solicitation. https://www.nsf. gov/pubs/2013/nsf13552/nsf13552.htm. Accessed 24 April, 2017.

Novak, G. M., Patterson, E. T., Gavrin, A. D., & Christian, W. (1999). *Just-in-time teaching: blending active learning with web technology*. Upper Saddle River, NJ: Prentice Hall.

Olson, S., & Riordan, D. G. (2012). *Report to the president: engage to excel: producing One million additional college graduates with degrees in science, technology, engineering, and mathematics Washington*. D.C.: The President's Council of Advisors on Science and Technology.

Onwuegbuzie, A. J., Slate, J. R., Leech, N. L., & Collins, K. M. T. (2007). Conducting mixed analyses: a general typology. *International Journal of Multiple Research Approaches, 1*, 4–17.

Osterman, K. F. (1990). Reflective practice: a new agenda for education. *Education and Urban Society, 22*(2), 133–152.

Pearson, R. (2016). GoodmanKruskal: Association analysis for categorical variables. R Package. https://CRAN.R-project.org/package=GoodmanKruskal. Accessed 10 March, 2017.

Penberthy, D. L., & Millar, S. B. (2002). The "hands-off " as a flawed approach to disseminating innovation: Lessons from chemistry. *Innovative Higher Education, 26*, 251–270.

Prince, M., Borrego, M., Henderson, C., Cutler, S., & Froyd, J. (2013). Use of research-based instructional strategies in core chemical engineering courses. *Chemical Engineering Education, 47*(1), 27–37.

Project Kaleidoscope. (2002). *Report on reports: recommendations for action in support of undergraduate science, technology, engineering and mathematics*. Washington, D.C.: Project Kaleidoscope.

Project Kaleidoscope. (2006). *Report on Reports II: Transforming America's Scientific and Technological Infrastructure: Recommendations for Urgent Action*. Washington, D.C.: Project Kaleidoscope.

QSR International Pty Ltd. (2011). *NVivo, Version 9*. Australia: Melbourne.

R Core Team (2016). R: A language and environment for statistical computing, Version 3.3.2. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Raker, J. R., & Holme, T. A. (2014). Investigating faculty familiarity with assessment terminology by applying cluster analysis to interpret survey data. *Journal of Chemical Education, 91*(8), 1145–1151.

Russell, T. (2006). Can reflective practice be taught? *Reflective Practice, 6*(2), 199–204.

Saldaña, J. (2013). *The Coding Manual for Qualitative Researchers* (2nd ed.). Thousand Oaks, CA: SAGE.

Schön, D. A. (1983). *The reflective practitioner: how professionals think in action*. New York: Basic Books.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: why undergraduates leave the sciences*. Boulder, Colorado: Westview Press.

Shadle, S. E., Marker, A., & Earl, B. (2017). Faculty drivers and barriers: laying the groundwork for undergraduate STEM education reform in academic departments. *International Journal of STEM Education, 4*(1), 8.

Slevin, J. F. (2001). Engaging intellectual work: the faculty's role in assessment. *College English, 63*(3), 288–305.

So, W. W. (1997). A study of teacher cognition in planning elementary science lessons. *Research in Science Education, 27*(1), 71–86.

Southerland, S. A., Nadelson, L., Sowell, S., Saka, Y., Kahveci, M., & Granger, E. M. (2012). Measuring one aspect of teachers' affective states: development of the science teachers' pedagogical discontentment scale. *School Science and Mathematics, 112*(8), 483–494.

Southerland, S. A., Sowell, S., Blanchard, M., & Granger, E. M. (2011a). Exploring the construct of pedagogical discontentment: A tool to understand science teachers' openness to reform. *Research in Science Education, 41*(3), 299–317.

Southerland, S. A., Sowell, S., & Enderle, P. (2011b). Science teachers' pedagogical discontentment: Its sources and potential for change. *Journal of Science Teacher Education, 22*(5), 437–457.

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., et al. (2018). Anatomy of STEM teaching in North American universities. *Science, 359*(6383), 1468–1470.

Stark, J. S. (2000). Planning introductory college courses: content, context and form. *Instructional Science, 28*(5), 413–438.

Stark, J. S., Lowther, M. A., Ryan, M. P., & Genthon, M. (1988). Faculty reflect on course planning. *Research in Higher Education, 29*(3), 219–240.

Swartz, L. B., Cole, M. T., & Shelley, D. J. (2010). Instructor satisfaction with teaching business law: Online vs. onground. *International Journal of Information and Communication Technology Education, 6*(1), 1–16.

Talanquer, V. (2014). DBER and STEM education reform: are we up to the challenge? *Journal of Research in Science Teaching, 51*(6), 809–819.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: combining qualitative and quantitative approaches. Applied Social Research Methods Series, No. 46*. Thousand Oaks, CA: Sage.

Teasdale, R., Viskupic, K., Bartley, J. K., McConnell, D., Manduca, C., Bruckner, M., et al. (2017). A multidimensional assessment of reformed teaching practice in geoscience classrooms. *Geosphere, 13*(2), 608–627.

Tomanek, D., Talanquer, V., & Novodvorsky, I. (2008). What do science teachers consider when selecting formative assessment tasks? *Journal of Research in Science Teaching, 45*(10), 1113–1130.

Turpen, C., & Finkelstein, N. D. (2009). Not all interactive engagement is the same: variations in physics professors' implementation of peer instruction. *Physical Review Special Topics-Physics Education Research, 5*(2), 020101.

Tyler, R. W. (1950). *Basic principles of curriculum development*. Chicago: University of Chicago Press.

Warfa, A.-R. M. (2017). Mixed-methods design in biology education research: approaches and uses. *CBE – Life Sciences Education, 15*(4), 1–11.

Wasilik, O., & Bolliger, D. U. (2009). Faculty satisfaction in the online environment: an institutional study. *Internet and Higher Education, 12*(3-4), 173–178.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Zieffler, A., Park, J., Garfield, J., Delmas, R., & Bjornsdottir, A. (2012). The statistics teaching inventory: A survey on statistics teachers' classroom practices and beliefs. *Journal of Statistics Education, 20*(1), 1–29.

Publisher's Note

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