

Exploring the Relationship between Teacher Knowledge and Active-Learning Implementation in Large College Biology Courses

Tessa C. Andrews,^{1*} Anna Jo J. Auerbach,² and Emily F. Grant³

¹Department of Genetics, University of Georgia, Athens, GA 30602; ²Department of Biological Sciences, Salisbury University, Salisbury, MD 21801; ³Department of Biology, Hastings College, Hastings, NE 68901

ABSTRACT

Not all instructors implement active-learning strategies in a way that maximizes student outcomes. One potential explanation for variation in active-learning effectiveness is variation in the teaching knowledge an instructor draws upon. Guided by theoretical frameworks of pedagogical content knowledge and pedagogical knowledge, this study investigated the teaching knowledge instructors used in planning, implementing, and reflecting on active-learning lessons in large courses. We used a preinstruction interview, video footage of a target class session, and a postinstruction interview with stimulated recall to elicit the teaching knowledge participants used. We then conducted qualitative content analysis to describe and contrast teaching knowledge employed by instructors implementing active learning that required students to generate their own understandings (i.e., generative instruction) and active learning largely focused on activity and recall (i.e., active instruction). Participants engaging in generative instruction exhibited teaching knowledge distinct from that of participants focused on activity. Those using generative instruction drew on pedagogical knowledge to design lessons focused on students generating reasoning; integrated pedagogical content knowledge and pedagogical knowledge to plan lessons to target student difficulties; and created opportunities to develop new pedagogical content knowledge while teaching. This work generated hypotheses about the teaching knowledge necessary for effective, generative active-learning instruction.

INTRODUCTION

Active-learning instruction can be much more effective at promoting the development of conceptual understanding and scientific thinking skills than traditional lectures (e.g., Freeman *et al.*, 2014). Active learning can also support the retention of students from groups historically underrepresented in science, technology, engineering, and mathematics (STEM) and reduce achievement gaps between students who have been underserved and more privileged students (e.g., Springer *et al.*, 1999; Haak *et al.*, 2011). However, active learning often falls short of this potential (e.g., Andrews *et al.*, 2011), most likely because it is implemented in ways that do not fully support student learning (Dancy *et al.*, 2016). Implementing active learning effectively makes different demands on instructors than delivering lectures and may require specialized knowledge of teaching and learning. Currently, we lack in-depth investigations of the teaching knowledge that instructors rely on to effectively implement active-learning instruction. This study contributes to filling that gap by taking an in-depth look at the relationship between teaching knowledge and teaching practices among active-learning instructors teaching large undergraduate biology courses.

Jennifer Knight, *Monitoring Editor*

Submitted Jan 14, 2019; Revised Jul 10, 2019;
Accepted Jul 24, 2019

CBE Life Sci Educ December 1, 2019 18:ar48
DOI:10.1187/cbe.19-01-0010

*Address correspondence to: Tessa Andrews
(tandrews@uga.edu).

© 2019 T. C. Andrews *et al.* CBE—Life Sciences Education © 2019 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

Guiding Theoretical Frameworks and Relevant Prior Research

The term “active learning” is widely used and poorly defined (e.g., Cooper, 2016). We use the ICAP (interactive, constructive, active, passive) framework to distinguish between engaging students in passive, active, and generative cognitive work (Chi and Wylie, 2014). Passive engagement occurs when students sit and listen, as is often the case in traditional lecture courses. Active work physically engages students and mostly requires recalling information. Generative work, which encompasses the interactive and constructive levels of ICAP, involves students working individually or collaboratively to generate ideas and products that go beyond what has been presented to them (Chi and Wylie, 2014). Generative work is expected to lead to deep understanding and the potential for transferring that understanding across contexts, while active work leads to shallow understanding with limited potential for transfer (Menekse *et al.*, 2013; Chi and Wylie, 2014). Thus, we are particularly interested in instruction that provides students with multiple opportunities for generative work during class.

Our research drew on two frameworks of teaching knowledge: pedagogical content knowledge (PCK) and pedagogical knowledge (PK). We will consider these as distinct knowledge bases for teaching, which is how they have been treated in prior theory and research (e.g., Gess-Newsome, 2015). Importantly, both knowledge bases go well beyond content knowledge of the discipline. Teaching knowledge is specific to the work of teaching and is not developed through training in a discipline alone. Furthermore, teaching experience is necessary, but often insufficient, to build teaching knowledge (e.g., Chan and Yung, 2018). We currently lack empirical work investigating PCK and PK together, how these knowledge bases inform teaching practices, and how they interact with each other. Prior work suggests that both knowledge bases are important for active-learning instruction in large college courses (e.g., Auerbach *et al.*, 2018), but professional development efforts would benefit from additional elucidation of how these knowledge bases are used in practice by instructors.

Scholarly debate persists regarding the nature of teaching knowledge and how it should be investigated. Some conceptualizations of teaching knowledge include collective knowledge generated through research and best practices and made available for others to use (e.g., Shulman, 1987). This view holds that instructors may develop knowledge that is useful to their teaching by reading literature, such as research literature about how

undergraduates commonly think about a topic (e.g., Smith *et al.*, 2016; Ziadie and Andrews, 2018). Other researchers view teaching knowledge as highly contextualized to the instructor, students, and context, and as manifested in the moment while teaching (e.g., Alonzo and Kim, 2016). They contend that teaching is fast and responsive and necessarily interweaves belief, knowledge, and skills (e.g., Petrou and Goulding, 2011; van Driel *et al.*, 2014). In this view, teaching knowledge cannot be meaningfully examined outside the context of teaching, nor is it readily shared among instructors. We allow for both conceptualizations of teaching knowledge in this work. We examine the knowledge instructors use before and following teaching, as well as knowledge manifested in real time while teaching.

Pedagogical Content Knowledge

PCK is topic-specific knowledge of teaching and learning used in planning, implementing, and reflecting on instruction for a particular grade level (Gess-Newsome, 2015). Instructors may need unique PCK for each topic they teach. For example, an instructor who uses PCK for teaching genetic drift will need additional PCK for teaching natural selection. We concentrate here on two components of PCK that have robust theoretical and empirical grounding (e.g., Park and Oliver, 2008; Depaepe *et al.*, 2013; Chan and Yung, 2015). The first component of PCK that we investigated was knowledge of student understanding (Table 1). PCK about student understanding allows instructors to anticipate what students are likely to think and what they will find confusing within a topic (Ball *et al.*, 2008). This knowledge is also called upon when instructors hear and make sense of student thinking in real time while teaching (Ball *et al.*, 2008). For example, PCK for teaching population genetics includes awareness that students often think that dominant alleles necessarily improve fitness and become more common in a population over time (Abraham *et al.*, 2014). This knowledge about student thinking may lead an instructor to teach the concept of dominance differently when discussing Mendelian genetics in order to avoid promoting ideas that will become problematic when students try to learn population genetics.

The second component of PCK investigated in this study was knowledge of instructional strategies and representations, which involves approaches to teaching a particular topic (Table 1). Instructors draw on this knowledge as they plan what work students will do during class to learn a specific topic and what representations will help or hinder student understanding. Instructors also use this knowledge in real time as they

TABLE 1. Overview of relevant knowledge bases for teaching and their constitutive components

Pedagogical content knowledge (PCK) is topic-specific knowledge of teaching and learning.

Knowledge of student understanding includes awareness of students' prior knowledge about a topic, conceptual difficulties around a topic, and common inaccurate ideas about a topic, and how student thinking about a topic is likely to change with instruction (e.g., Magnusson *et al.*, 1999; Park and Oliver, 2008).

Knowledge of instructional strategies and representations includes awareness of effective examples, analogies, problems, activities, case studies, and visual representations that make a specific topic accessible to students and facilitate learning (e.g., Shulman, 1987; Magnusson *et al.*, 1999).

Pedagogical knowledge (PK) is generalizable knowledge of teaching and learning.

Knowledge of creating opportunities for generative work includes awareness that students learn from cognitively engaging in challenging work during class, knowing the types of tasks and problems that require generative work, and understanding how to facilitate in class to maintain cognitive engagement (e.g., Auerbach and Andrews, 2018).

Knowledge of monitoring and responding to student thinking includes awareness of approaches to elicit detailed student thinking during class and the purposes of accessing student thinking and altering instruction accordingly (e.g., Auerbach and Andrews, 2018).

decide how to proceed in a lesson based on student contributions, including deciding to pose a new question, return to a prior question, or take a new approach to engaging students with a topic (Ball *et al.*, 2008). For example, PCK for teaching phylogenetic trees includes awareness that students are more successful at reading branching cladograms than those in a ladder style (e.g., diagonal) and therefore benefit from first learning to read branching cladograms (Novick and Catley, 2013).

Though theory proposes that PCK is crucial to effective teaching at all levels, there are few studies of PCK among college instructors. Three in-depth case studies with individual mathematics instructors shed some light on PCK needed for evidence-based instruction. Lack of PCK hindered these instructors as they implemented inquiry-oriented curriculum. Specifically, instructors struggled to facilitate class discussions, because they could not anticipate the difficulties students would encounter as they learned particular concepts, nor the likely consequences of these difficulties as students engaged in particular learning activities (Speer and Wagner, 2009). Additionally, they did not predict how students would perceive examples, counterexamples, and explanations about a topic. These knowledge deficits made it harder to make sense of student thinking in real time during class and to use students' ideas to further the lesson progression, especially when the ideas were vaguely expressed or partially formed (Wagner *et al.*, 2007; Speer and Wagner, 2009; Johnson and Larsen, 2012). Additionally, lack of awareness of the kind and quality of understandings students could be expected to have about particular topics made it hard to make decisions about when to move from one topic to the next (Wagner *et al.*, 2007). Another study of 44 college biology instructors revealed that expert active-learning instructors displayed more PCK as they analyzed videos of lessons than did novice active-learning instructors (Auerbach *et al.*, 2018). Together, these studies suggest an important role for PCK in active-learning instruction in STEM courses.

Pedagogical Knowledge

PK is generalizable knowledge of teaching and learning. We expect PK, unlike PCK, to be transferable across different topics an instructor teaches. PK has received considerably less attention from researchers than PCK. As a result, there is sparse empirical and theoretical work to define components of PK. Auerbach and Andrews (2018) proposed a framework of components of PK relevant to using active-learning instruction in large undergraduate biology courses. This framework includes seven interrelated components that emerged from qualitative analysis of the thinking of 77 instructors using active-learning strategies. Instructors' PK included knowledge related to principles of how people learn, student motivation, classroom environment, general instructional approaches, lesson structure and design, and classroom management. These components have considerable overlap with other definitions of pedagogical knowledge from work with K–12 instructors (e.g., Grossman and Richert, 1988; Morine-Dersheimer and Kent, 1999; König *et al.*, 2014).

We concentrate here on two components of PK, because there is some evidence this knowledge is characteristic of expert active-learning instructors who teach large courses (Auerbach *et al.*, 2018). Knowledge of creating opportunities for generative work encompasses constructivist and social constructivist ideas about how students learn (Table 1; Auerbach and Andrews,

2018). For example, instructors drawing on this PK can create opportunities for students to use logic and develop their own reasoning during class and can resist providing students with answers in favor of asking follow-up questions (Auerbach and Andrews, 2018).

The second component of PK examined in this study is knowledge of monitoring and responding to student thinking. This component deals with approaches to accessing student thinking during class in order to learn where students were struggling, gauge the effectiveness of a task, and respond in real time to student thinking (Table 1; Auerbach and Andrews, 2018). This component of PK may interact with knowledge of creating opportunities for generative work, because generative work shows the instructor what students are thinking. Additionally, the chance to reason about student thinking can help instructors make real-time modifications to support students in generative work (Auerbach and Andrews, 2018). Instructors drawing on this component of PK deliberately seek out student thinking during class by talking to students while they work, asking open-ended questions, and asking students to share their thinking with the whole class.

Research on instructor practice and decision making in the classroom can shed light on important PK for active-learning instruction. Seidel and colleagues (2015) studied the language used by instructors that was not directly related to content. This “instructor talk” could provide evidence of the PK that instructors rely on while teaching. In fact, the emergent categories of instructor talk identified by Seidel and colleagues (2015) have substantial overlap with the framework of PK proposed by Auerbach and Andrews (2018). Specifically, instructor talk observed in an active-learning biology course cotaught by two instructors related to opportunities for generative work, monitoring and responding to student thinking, motivating students, equity, metacognition, and managing the logistics of an active-learning lesson (Seidel *et al.*, 2015). For example, one category of instructor talk is described by researchers as “setting the tone for the course as a whole or for specific activities within the course.” Instructor quotes from this category prompt students to share their reasoning with peers and to support their answers with evidence (Seidel *et al.*, 2015). This aligns with knowledge of creating opportunities for generative work. Collectively, we have only just begun to investigate the specific PK that instructors rely on to effectively plan and implement active-learning instruction.

Research Aims and Contributions

The goal of this research was to describe how college biology instructors use components of PCK and PK to design and implement opportunities for students to engage in generative work in large classes. This builds on prior research by examining how instructors use teaching knowledge *in their own practice*. We analyzed teaching practices to identify instructors who created frequent opportunities for students to engage in generative work during class. We then qualitatively contrasted teaching knowledge used by these instructors with teaching knowledge used by instructors who primarily create opportunities for activity and recall.

This in-depth exploratory study makes novel contributions. We examine both PCK and PK used by college instructors in their own teaching. Prior studies of the knowledge used by

college STEM instructors while teaching have generally examined just one or two instructors and focused only on PCK (e.g., Wagner *et al.*, 2007; Speer and Wagner, 2009; Johnson and Larsen, 2012). We take a comprehensive and highly contextualized approach by examining knowledge used while planning, implementing, and reflecting on teaching (Chan and Hume, 2019). This captures knowledge enacted and embedded in an instructor's practice, as well as an instructor's capacity to reason through instructional decisions, providing a deeper look at teacher knowledge than approaches using pencil-and-paper tests, surveys, teacher reflections, lesson plans, or classroom video alone (Chan and Hume, 2019).

METHODS

Participants

We aimed to recruit participants who self-identified and were seen by colleagues as active-learning instructors and who taught large (50+ students) undergraduate courses in the life sciences. We expected that these instructors would vary in the degree to which they engaged students in active versus generative cognitive engagement. Therefore, we asked colleagues to recommend instructors in life sciences departments who used active learning, and we contacted those instructors, seeking a sample of instructors across a broad experience range.

Our final sample included 13 life sciences instructors from five large, research-intensive institutions in the Southeastern United States (Supplemental Table S1). We studied 12 of these instructors in the context of introductory biology courses for biology majors, non-biology majors, or a mix of majors and nonmajors. We investigated one participant in the context of an upper-division course for life sciences majors. The mean number of students in participants' courses was 167 (SD = 72). Participants had taught college courses for a median of 11 semesters and ranged in prior experience from 1 to 42 semesters. All but one participant reported having previously participated in 40 or more hours of teaching professional development and seven had previously led teaching professional development. All instructors had long-term faculty or staff appointments at their institutions, and two were on a tenure-track. More than half (62%) of the participants identified as female, two (15%) identified with races other than "white," and none identified as Hispanic. All research was approved by the University of Georgia Institutional Review Board under protocol ID #00004989.

Eliciting Teacher Knowledge

We aimed to elicit teacher knowledge employed in the complex tasks of teaching within the context of an instructor's own classroom. Our approach captured participants' thinking as they planned, implemented, and reflected on instruction. In keeping with our guiding theoretical frameworks of teacher knowledge, our interviews elicited PCK and PK.

Our overall approach included a preinstruction interview, filming a "target class" session and creating short video clips, and a postinstruction stimulated recall interview. We also collected teaching materials from each participant, including the course syllabus and calendar, the exam covering the target class, PowerPoint slides, and the questions or problems that students were asked to complete during the target lesson.

The objective of preinstruction interviews was to elicit knowledge used in planning the target class. We asked instructors to share lesson materials before the interview, including preclass work, in-class work, and slides. We used a semistructured interview that probed the aims and design of preclass work, learning objectives for lesson, how progress toward objectives would be monitored, students' prior knowledge and anticipated difficulties with topics, work students would do during class, how the instructor communicated with students about the rationale for in-class work, and what motivated students to work (see full interview protocol in Appendix A in the Supplemental Material). Interview questions specifically probed instructors' rationales for their thinking and decisions. This interview lasted ~60 minutes and occurred 1 to 2 days before the target class. It was often conducted virtually using a video-conferencing platform.

We video-recorded the target class and created video clips as stimuli to help the instructors recall what they were thinking in "real time" while teaching. We filmed from the back of the classroom and captured instructor behavior, student behavior, and high-quality audio from the instructor and nearby students. We identified a total of ~10 minutes of video footage for each instructor, divided among short clips to use as stimuli in the postinstruction interview. We used five criteria to select segments of the target class to use as interview stimuli: at least 1) three clips when the instructor appeared to be making decisions in the moment; 2) three clips in which the instructor had access to student thinking; 3) one clip showing students working individually or in small groups; 4) one clip with students sharing their ideas with the class; and 5) two clips showing the instructor interacting with students in a way that was not about content. Each clip met multiple criteria. Using specific criteria to identify clips ensured that all participants were asked about similar aspects of active-learning instruction in the postinstruction interview, even though their teaching varied. We met each of these criteria for all participants.

The objective of the postinstruction interview was to facilitate the recall of knowledge used during instruction, which is often tacit, and to elicit knowledge drawn on while reflecting on the target class. Following the methods of Alonzo and Kim (2016), the participant and interviewer watched the selected clips of classroom footage, and the interviewer asked the participant to explain anything he or she could recall thinking in that teaching moment. Watching video prompts detailed recall beyond what a participant would otherwise remember. Next, we asked predetermined questions about the videos to further elucidate teaching knowledge. We could not use a strict semistructured interview protocol, because the events in the videos varied across participants. However, we still needed to probe participants in similar ways to make comparisons across instructors. Therefore, we developed a collection of interview questions related to the specific components of PCK and PK we aimed to study (see full list of questions in Appendix B in the Supplemental Material). We then posed a minimum number of questions related to each component to each participant. For example, to elicit thinking about creating opportunities for generative work (PK), we could ask, "What were you thinking when you directed students to explain their reasoning to a neighbor?" To elicit knowledge of student understanding (PCK), we could ask,

“Do you have any insight into why that concept might be difficult for students?” The postinstruction interview lasted ~60 minutes and occurred in person within 48 hours of the target class to maximize participants’ recall.

Analysis of Teaching Practices

We used two quantitative analyses of teaching practices to distinguish between participants who taught lessons with frequent opportunities for generative work and participants who created few opportunities for generative work.

One analysis of instructional practices consisted of two simple metrics to quantify opportunities for generative work. We determined the proportion of class time that students actively worked. This included time spent working individually and time spent collaborating in groups. It did not include time when most students were only listening, such as instructor introductions to an activity or whole-class discussion. This is similar to a dimension of the Practical Observation Rubric to Assess Active Learning (PORTAAL) that quantifies the amount of class time that students had the chance talk through content in class (element P1; Eddy *et al.*, 2015). This quantification of teaching practice alone is insufficient because it does not distinguish active and generative work.

Therefore, we also determined the proportion of class time during which students were asked to engage in generative work. We operationalized generative work as students engaging in tasks requiring higher-order cognitive skills (HOCs). We identified each question or problem students answered during class using PowerPoint slides, in-class materials, and video recordings. We also used these in-class materials and preclass materials to determine the context in which the instructor asked each question. This is important, because the same question can require HOCs in one context and only lower-order cognitive skills (LOCs) in another context. For example, a question can require students to apply their understanding to a novel context if they have not encountered the context previously. The same question will require students to recall information only if they have encountered the context of the question previously.

We used the method and resources for determining Bloom’s level described by Crowe *et al.* (2008), as well as training materials provided by Eddy *et al.* (2015). We categorized each question or problem posed to students as requiring HOCs or only LOCs (Crowe *et al.*, 2008). Questions requiring only LOCs ask students to remember facts and equations, identify parts, describe a term or process in their own words, or provide an example. This corresponds to questions at the knowledge, comprehension, and lower application levels described by Crowe *et al.* (2008). Questions requiring HOCs ask more of students, such as using several pieces of information to predict an outcome, interpreting data in a graph with multiple variables and explaining the biological implications, designing an experiment to answer a specific question, creating a cladogram using data on a variety of traits, or using a pedigree to develop a hypothesis for how a disease is transmitted (Crowe *et al.*, 2008). This corresponds to questions at the higher application, analysis, synthesis, and evaluation levels described by Crowe *et al.* (2008).

We defined higher application questions as those requiring multiple steps to solve. Additionally, students had to determine and carry out those steps themselves, rather than having the

instructor explain the steps before students encountered the problem. This is similar to how Arneson and Offerdahl (2018, p. 3) define higher application questions: “tasks that involve conceptual understanding and require selection, modification, or manufacture of a procedure to fit the situation.” In contrast, application questions that entail only one step to solve or multiple steps previously laid out by the instructor require only LOCs.

We determined the proportion of class time in which students were engaged in HOCs and LOCs. In some cases, students worked on a series of questions in one time period, making it unclear how much time was spent on each individual question. In those cases, we assumed that questions requiring HOCs would take twice as much time as questions requiring LOCs and divided the full time period across the questions. We took this approach in order to err on the side of assuming that instructors dedicated more time to HOCs. Two researchers independently rated each question posed to students in a target lesson as requiring only LOCs or also HOCs, while carefully considering the context of the question. Our percent agreement for 133 questions was 72%. We discussed any disagreements until we reached consensus about the cognitive demands placed on students for each question.

Our second quantitative analysis of teaching practices used an established protocol that allows comparison to other data sets. We analyzed each lesson using the Classroom Observation Protocol for Undergraduate STEM (COPUS; Smith *et al.*, 2013). This protocol divides the lesson into 2-minute segments. Researchers code each 2-minute segment to indicate the instructor and student behaviors observed. After training to use the COPUS as a team, A.J.J.A. and E.F.G. coded four (30.7%) lessons independently and achieved an interrater reliability of 0.85. We calculated interrater reliability using Fleiss’s kappa, which assumes raters were chosen at random and accounts for agreement by chance (Nelson and Edwards, 2015). E.F.G. coded the remaining lessons independently. The data generated by the COPUS are the proportion of 2-minute segments of class during which 25 different behaviors occurred.

We then used the COPUSAnalyzer to create a meaningful synthesis of these data. The COPUSAnalyzer uses four instructor behaviors and four student behaviors to categorize a lesson into one of seven COPUS profiles, each representing a unique instructional profile (Stains *et al.*, 2018). The lessons taught by our participants fell into four different profiles, COPUS profiles 3, 4, 5, and 7. COPUSAnalyzer creators refer to COPUS profiles 5 and 7 as “student-centered.” Profile 5 represents instruction that uses a variety of group-work strategies consistently, and profile 7 represents a similar variety of group-work strategies, but with less consistent usage (Stains *et al.*, 2018). Profiles 3 and 4 are labeled as “interactive lecture,” and include lessons in which lecture is supplemented with group work. Profile 3 involves group work that is not associated with clickers, and Profile 4 involves clicker questions with group work. According to the reports provided by the COPUSAnalyzer, the algorithm is able to predict cluster membership correctly more often for some profiles than for others. Specifically, it correctly predicts Profile 3 membership 86.2% of the time, profile 4 membership 96% of the time, profile 5 membership 77.6% of the time, and profile 7 membership 58.8% of the time.

After categorizing participants as exhibiting generative instruction or active instruction based on a triangulation of

these analyses of teaching practices, we used Welch's unequal variance *t* tests to compare the proportion of time that students worked and the proportion of time spent on HOCs in these two types of instruction. We also calculated an effect size for these comparisons to estimate the magnitude of the difference in how student time was spent in classes exhibiting generative versus active instruction. We calculated effect size as Hedge's *g*, a modification of Cohen's *d* that weights group's standard deviations by their size (*n*) and removes a small positive bias that affects Cohen's *d* calculations in the case of small sample sizes (Maher *et al.*, 2013).

Qualitative Analysis of Teaching Knowledge

Our qualitative analysis of teaching knowledge aimed to characterize how instructors drew on two components of PCK and two components of PK in their planning, implementation, and reflection on the target lesson, and how they integrated knowledge components. We wanted to identify teaching knowledge that was important to generative active-learning instruction, so we analyzed the teaching knowledge of each participant and then contrasted those who exhibited generative instruction and those who primarily exhibited active instruction. We examined how instructors integrated components within PK and within PCK, such as drawing on both knowledge of student understanding and knowledge of instructional strategies to write in-class questions. We also examined how instructors integrated PK with PCK, such as using knowledge of student understanding of a topic and knowledge of monitoring and responding to student thinking to circulate around the classroom, check in with students, and ask prompting questions about specific topics expected to be difficult for students. The final outcome of our analyses was three themes characterizing teaching knowledge that was common across instructors exhibiting generative instruction. We arrived at these themes by engaging in highly collaborative and iterative constant comparative analysis to characterize the teacher knowledge exhibited and by holistically contrasting the thinking behind generative and active instruction to identify teaching knowledge that may facilitate generative instruction.

Pre- and postinstruction interviews formed the raw data for analyzing instructor use of PK and PCK, and other data sources provided important context for making sense of participant thinking. In addition to reading each interview multiple times, researchers watched the video clips used as stimuli in the postinstruction interview and reviewed video and audio footage of the full target class and class materials.

We engaged in constant comparative analysis to elucidate themes in our data. Constant comparative analysis involves iteratively and continually making comparisons within your data (Glaser and Strauss, 1967; Charmaz, 2006; Birks and Mills, 2011). Our initial qualitative analysis involved line-by-line coding of interview transcripts to identify any evidence of PK and PCK. We started with a codebook grounded in theory that included each knowledge component stipulated in the PCK and PK frameworks. We soon expanded this codebook to capture the ways in which components of PK and PCK were evident in instructors' thinking while planning, implementing, and reflecting on instruction. Two researchers (T.C.A. and A.J.J.A.) coded full transcripts independently and then compared and discussed the coding to reach consensus. We retained only

codes that we agreed on. We refined the descriptions of codes as we analyzed more data. We then returned to transcripts previously coded and reanalyzed them, again working independently and then discussing any disagreements to reach consensus. We also read all quotes within each code. This allowed us to compare quotes within codes to ensure that they expressed the same ideas and did not overlap with other codes. We repeated this process several times throughout our analysis.

We approached qualitative analysis in two waves, first analyzing transcripts for evidence of PCK and then analyzing them for evidence of PK. This resulted in repeatedly reading the same transcripts for different purposes, which further contributed to constant comparison and started the process of axial coding. Axial coding involves relating codes to one another and grouping the codes into larger categories (Charmaz, 2006). It also helps determine which codes are most dominant or important to addressing the research questions and which are more peripheral (Saldaña, 2013). At this point in the analysis, we began to detect evidence of knowledge integration. As codes emerged that captured teaching knowledge about topic-specific instructional strategies (i.e., PCK), we recognized that some focused on passive instructional approaches (e.g., lecture) and some focused on generative instructional approaches (e.g., present students with an incomplete model that they work to construct). We were able to further refine the organization of PCK codes as we made progress in analyzing transcripts for evidence of PK.

We also analyzed each participant's teaching knowledge holistically and contrasted the knowledge displayed by those who engaged in generative versus primarily active instruction. This is another example of constant comparison in our analysis. We wrote detailed analytic memos describing the PCK and PK used by each instructor and any evidence of integration among knowledge components. These memos discussed the codes applied throughout the pre- and postinstruction interview transcripts for each instructor and also reflected what could be learned by comparing participants with one another. Memos noted the knowledge instructors used as well as knowledge that was absent. Memo writing engages researchers in synthesizing their analysis of the data across codes, across data for a single participant, and across multiple participants (Charmaz, 2006). It leads to new insights and begins to uncover major themes in the data (Saldaña, 2013). The memo-writing process was also iterative and reciprocal with the coding process. As we recognized something new about the use of teaching knowledge in one participant, we returned to previously analyzed data and reread transcripts for evidence of the same pattern.

Our qualitative analysis resulted in the three themes that are the main results of this study. These themes represent key qualities of PK and PCK underlying generative instruction that were much less commonly exhibited by instructors exhibiting active instruction. After arriving at these themes, we returned to the data to seek evidence of negative cases. Negative cases are exceptions to patterns or themes in the data (Patton, 1990). Searching for and analyzing negative cases can help researchers appreciate variation in the data and can result in further refinement of emergent themes (Cresswell and Miller, 2000; Charmaz, 2006). We searched for two types of negative cases: 1) participants exhibiting generative instruction who were not aligned with a theme, and 2) participants exhibiting active instruction who were aligned with a theme. We examined code

frequencies and reanalyzed full transcripts in the pursuit of negative cases. One participant represented a negative case for two of the themes. Her target lesson primarily required active work, but the knowledge she displayed was more similar to that of instructors engaging students in generative work. We discuss her knowledge and practice specifically under the two themes for which she was a negative case.

Trustworthiness of Our Qualitative Approach

The trustworthiness (i.e., rigor) of qualitative data analysis stems from its credibility, transferability, dependability, and confirmability (Guba, 1981). We discuss these four criteria of trustworthiness in relation to our analytic approach.

Credibility deals with the degree to which a study measures what is intended, which makes it similar to internal validity (e.g., Shenton, 2004). Several attributes of our qualitative analysis contribute to its credibility. First, we carefully designed our data collection to provide access to knowledge applied throughout the stages of teaching. Our approach built on established methods of accessing instructor knowledge (e.g., Alonzo and Kim, 2016), increasing confidence that we are measuring the construct of interest. Second, we were able to triangulate data about instructor knowledge displayed in interviews before and after teaching, as well as instructor knowledge displayed through their instructional practices. Third, we have been highly transparent regarding our research methods, allowing the reader to carefully consider how the research process might have shaped the data collected and the analysis. Fourth, we have undertaken negative case analysis as a verification step for the key emergent themes. Finally, we describe here our positionality to this research, as we are integral parts of the qualitative analysis process. Authors T.C.A. and A.J.J.A. collaborated to complete all of the qualitative analyses. We both have experience teaching large biology courses and aim to design and implement generative active-learning instruction in our own teaching. We had taught some, but not all, of the topics taught in target lessons of our participants. We drew on our own teaching experiences to interpret data, but were careful to push back against our own assumptions in data collection and analysis. We asked follow-up questions to clarify what participants meant, even when we thought we understood their reasoning. We also considered the entirety of a participant's interview and observation data as we interpreted the meaning behind his or her words, and put more trust in ideas that appeared multiple times in a participant's data.

Transferability is concerned with the degree to which findings of one study have relevance to contexts beyond those specifically studied (e.g., Shenton, 2004). We recruited participants with a range of teaching experience from five different institutions who displayed a range of teaching practices. We are drawing on this sample to generate hypotheses to be tested in future studies, rather than to draw generalizable conclusions.

Dependability and confirmability deal with the degree to which interpretations of the data align with raw data and could be considered stable and repeatable across researchers, making them similar to reliability and objectivity, respectively (Anfara *et al.*, 2002; Shenton, 2004). Our reliance on constant comparison contributes significantly to the dependability and confirmability of our qualitative analysis. Two researchers became fully immersed in all of the data, so no one person was responsible

for interpreting any data. We also repeatedly engaged in both independent analysis and sense-making achieved through rigorous discussion of ideas. This approach leverages the coconstruction of ideas and guards against power dynamics in a researcher–researcher relationship that could influence analysis. Requiring consensus ensures that no single person's perspective can pervade the interpretation of data. In addition to constant comparison between researchers, we reanalyzed each transcript and coded segment multiple times, which contributes to reliability across time.

RESULTS

Instructional Practices

Instructional practices varied considerably among participants, even though each self-identified as an active-learning instructor and was recommended by colleagues as using active learning. Our triangulation of two analyses of instructional practices indicated that five participants engaged students in generative work during class and eight participants primarily engaged students in active work. Five instructors stood out as teaching lessons in which a greater proportion of class time was spent with students using HOCs, as well as a greater proportion of class time during which students worked (Figure 1). Analysis of teaching practices using the COPUSAnalyzer placed these same five instructors in COPUS profile 5 or 7, which indicates student-centered instruction. In contrast, the other eight instructors aligned with COPUS profiles 3 or 4, which indicate interactive lecture (Figure 1).

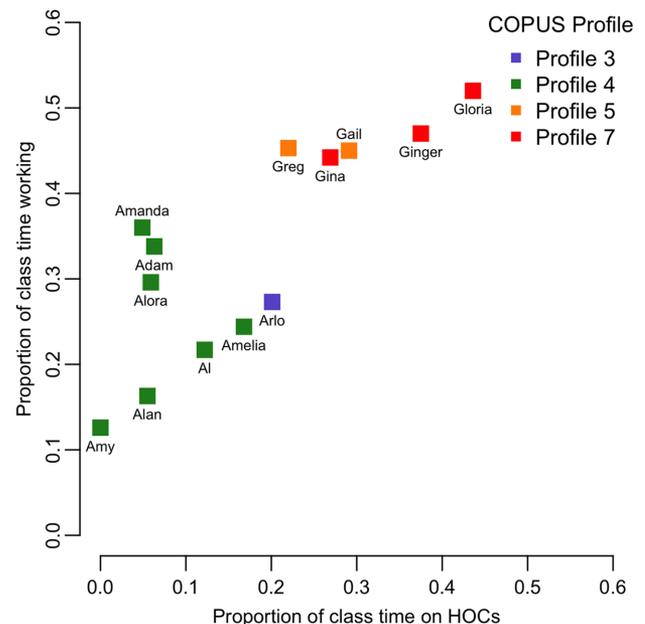


FIGURE 1. Generative instructors are different from active instructors in COPUS instructional profiles, proportion of class time spent working, and proportion of class time spent on higher-order cognitive skills (HOCs). Pseudonyms starting with the letter “G” designate “generative instructors” and pseudonyms starting with “A” designate “active instructors.” It is important to note that categorizations of generative and active were based on a single class period.

TABLE 2. Characterizations of how instructors asked students to spend class time, by generative and active instructors

	All instructors	Generative instructors	Active instructors
Mean proportion of class time spent working (SD)	0.34 (0.13)	0.47 (0.03)	0.22 (0.08)
Mean proportion of class time spent on HOCs (SD)	0.18 (0.14)	0.32 (0.07)	0.09 (0.08)

After dividing instructors into two groups based on the triangulation of these analyses, we compared their practices to determine whether they were statistically significantly different from each other. These results should be interpreted with caution, given our small sample size. Instructors exhibiting generative active learning taught lessons that engaged students in a significantly greater proportion of work during class time ($t = 6.70, p < 0.0001$). The effect size for this comparison was 3.6 (95% confidence interval: 1.6–5.6). This is considered a very large effect (Maher *et al.*, 2013), and it is interpreted to indicate that the mean time students spent working during class was 3.6 SDs higher in courses exhibiting generative instruction than those exhibiting active instruction.

Students in classes using generative instruction also spent a significantly greater proportion of class time on questions that required HOCs ($t = 5.03, p = 0.001$). The effect size for this comparison was 2.7 (95% confidence interval: 1.0–4.4), which is also considered a very large effect (Maher *et al.*, 2013). Students taught by these instructors had opportunities to engage in HOCs for a mean of 16 or 24 minutes for a 50- or 75-minute course, respectively. In contrast, students taught by instructors relying primarily on active work had opportunities to engage in HOCs for a mean of just 5 or 7 minutes in a 50- or 75-minute class, respectively (Table 2). In contrast, students taught by instructors relying primarily on active work had opportunities to engage in HOCs for a mean of just 5–7 minutes. Importantly, these two groups of instructors did not differ in terms of teaching experience ($t = 0.1114, p = 0.91$).

Teaching Knowledge

We were particularly interested in better understanding teaching knowledge used by instructors who create opportunities for generative work. Specifically, we describe three ways in which these instructors relied on teaching knowledge as they planned, implemented, and reflected on instruction. These three themes distinguished those who exhibited generative instruction from those exhibiting primarily active instruction in the target lesson. We dedicate a section below to each of these themes. At the end of each section, we describe how knowledge used by instructors displaying active instruction differed from knowledge used by those displaying generative instruction. It is important to recognize that the teaching knowledge used by all generative instructors was not identical, nor was the knowledge employed by active instructors. Additionally, instructors drew to a varying extent on components of PCK and PK that are not the focus of this paper.

The three themes emerging from our investigation are that generative instructors

1. drew on PK to teach lessons focused on students generating reasoning;
2. integrated PCK and PK to plan lessons to target common difficulties; and

3. used PK while teaching, creating opportunities to develop new PCK.

We depend heavily on quotes to illustrate our participants' thinking. Quotes are lightly edited for grammar, such as subject–verb agreement. We also removed phrases such as “like,” “um,” “sort of,” and “kind of” when they were used as filler words. For simplicity, we refer to participants as “generative” or “active” instructors throughout the results. It is important to note, however, that we are making these categorizations based on a single class session. All participant names have been replaced with pseudonyms. Pseudonyms for those exhibiting generative and active instruction begin with the letters “G” and “A,” respectively.

Generative Instructors Drew on PK to Teach Lessons Focused on Students Generating Reasoning

Generative instructors explained that students learn by generating explanations supported with reasoning. Therefore, they planned and implemented lessons that prioritized time for students to articulate their reasoning about biological concepts. This demonstrates knowledge of creating opportunities for generative work, a component of PK. Gail stated, “The more I teach, the more I think what I say is not very important.” She explained that “the chance to explain what they’re thinking ... is where the learning occurs.” She observed that generating explanations helps students realize what they do and do not understand:

If they’re just listening then they—I think when students listen, things make sense to them. They either zone out or they think, “Yeah. It makes a lot of sense.” But then when they have to explain something new in terms of what they think they know, they have to recall stuff and they have to put it in a correct order and they have to draw the correct relationships. And that takes work and the discussions give them that opportunity to do that ... I love it when they start to argue with each other because then they’re running up against some misconception that somebody has. And they really have to justify why they think a certain way.

Gina asked students to explain their reasoning when they answered questions, because she reasoned that it would lead to learning that would be more useful to students later in life. She wanted students to get away from “memorizing factoids” to “understanding how those factoids lead you to an answer.” She explained how students learn by explaining their reasoning to each other:

I think that, you’ll never know something so well until you have to teach it to someone else. So, when they’re talking in groups and pairs, they’re teaching each other. Maybe one gets it, the other one hasn’t gotten it yet, but then explaining it to each other really is the best way to understand something.

Generative instructors were all successful at getting the majority of students to talk to their peers during class, which was likely crucial to their efforts to support students in articulating their reasoning in such large classes. In each of their courses, the vast majority of students actively worked with their groups when they were asked to do so. This was notable, because we also observed courses where students were encouraged to work with their peers, but many opted not to do so. Some generative instructors used assigned groups of three to five students who sat in designated seats within the classroom and were stable throughout the semester. Other generative instructors did not have formalized groups, but engaged students in group work in each lesson, so students had assembled groups with whom they readily worked.

Generative instructors expected that developing and articulating reasoning would be challenging for students and therefore provided explicit verbal instructions to guide students. For example, Gloria started talking to students about how she wanted them to approach questions early in the semester:

And I try and say to them early in the semester too, you don't just turn to your neighbor and be like "I got C. Yep. I got C," that you have to be able to justify it because to be able to learn to apply things, you have to have the dialogue with yourself.

Gloria explained that "it is not a given that they will do that if you don't say that" and that it is important to their learning, because they may choose the right answer for the wrong reasons if they do not have "rich conversation with each other to make sure they were getting it right for the right reasons."

In addition to providing explicit verbal instructions to the whole class about supporting their answers with reasoning, generative instructors approached their interactions with small groups as a facilitator prompting for reasoning. They explained that this was important, because students could often answer a question but struggle to provide a rationale for that answer. Ginger described her thinking when a student asked her whether she should include a component in a vector. The component the student proposed was unnecessary for the biological problem Ginger had asked students to solve. Ginger responded to the student by asking, "Why would you do that?" In the postinstruction interview, she explained the rationale behind her response to the student:

And so asking her like, 'Why would you do that?' is getting her to think about the significance of that certain component in the context of what we were doing, which while it wouldn't have been harmful, it really wouldn't have been helpful in any way ... I could've just went off and said that. But yeah, I wanted to challenge her to tell me because she came up with that idea. I wanted to challenge her to explain to me, tell me where that came from, either where the idea came from or instead of just talking about a component that she was familiar with, I wanted to get at does she really understand the function of that component? So that then she can see why it wouldn't be important in this context. I think you could see and it felt like I could tell when I was interacting with her that she really got to thinking when I said, "Why would you do that?" So, if she had to give her rationale, it's where she kind of stopped like, "Whoa, I don't really know why." I feel like those are really important interactions.

One instructor who primarily engaged students in active work, Amanda, represents a negative case for this theme. She displayed knowledge of creating opportunities for generative work in interviews, but her instruction did not engage students in much generative work. Like instructors who engaged students in generative work, Amanda aimed for students to explain their reasoning when responding to questions in class. She prompted the class to do so and also carefully chose what she said to small groups who were working on a question:

What I was trying to do with these two exercises was students will often, it seems, you know, like "Okay, that's the correct answer. I'm going to pick that correct answer." But part of picking that correct answer then is also well, "Can I explain why I'm rejecting these other ones, to be certain about it?" And so I was trying to go through more prompting and teasing out what it was that they were thinking.

Students needed to articulate limited reasoning to fully answer the questions that Amanda posed in class, because these questions tended to require only comprehension or low-level application (Crowe *et al.*, 2008). In Amanda's case, knowledge about creating opportunities for generative work was insufficient to result in instruction that created ample opportunities for students to engage in generative work.

What else needed to be true for Amanda to design a lesson that engaged students in generative work? One way in which Amanda differed from generative instructors is that she indicated that the topics covered in the target lesson "tend to come pretty easily to my students." When asked how she might revise the lesson in the future, Amanda did not indicate that she would make the lesson more challenging or more focused on specific difficulties. In contrast, some of the quotes from generative instructors in the next theme highlight how they focused lessons on the topics that were particularly challenging for students.

Most instructors who displayed active instruction explained that they wanted students working during class, but they did not emphasize students generating reasoning. For some active instructors, the main goal of in-class work was to provide practice with the types of questions students would encounter on exams. This was an important perspective for Amelia. She designed much of her course to avoid "a complete disconnect between what you are practicing as a learner and then what you are expected to deliver as a learner at the end." She wanted students to get the chance to practice exam-like questions in a "safer" environment. Therefore, she ended each class with a series of multiple-choice questions. These questions represented the vast majority of student work time in her target lesson. She explained that she asked students to spend their time this way because "performance" in her course was measured by answering "hard multiple-choice questions on the exam," and the goal of in-class work was to practice for the performance evaluations. Similarly, Arlo included in-class quizzes because they provided students with "a lower-stakes environment to mess up." Some generative instructors also gave summative assessments composed of multiple-choice questions, yet their goals for in-class work focused on developing reasoning.

Some active instructors displayed ideas that seemed to value lecture as central to student learning. We did not observe these

ideas among generative instructors. Some active instructors thought that student work was important, because it would “break up the talking” and therefore keep students from “losing interest” or going “off task.” Implicit in this perspective is the idea that lecturing is effective and can be optimized with breaks to maintain students’ attention. Several active instructors considered clarifying, explaining, and transmitting information to be key instructor roles. Alan explained that he “is [in class] to answer questions and walk them through concepts,” and Al described his job as “weaving it all together in a narrative so that everyone can follow along.” These ideas about student learning are in stark contrast to those espoused by generative instructors. Alan and Al seem to be thinking that students learn well by listening. Generative instructors did not disregard the role of instructor explanations in student learning. In fact, up to 50% of their class time was spent with students listening rather than working. However, they also thought that formulating explanations supported with reasoning was key to the student learning they hoped to foster.

Generative Instructors Integrated PCK and PK to Plan Lessons to Target Common Difficulties

Generative instructors anticipated their students’ thinking and made it central to their lesson design. These instructors aimed for in-class work to elicit students’ thinking about a topic, and especially student thinking about difficult parts of a topic. This is evidence of the integration of two PCK components: knowledge of student understanding and knowledge of instructional strategies. Additionally, these instructors developed or selected instructional approaches that required students to do generative work, demonstrating integration of PK with PCK to plan lessons. Gloria explained that she was “trying to go to the places that I think are most challenging for them.” She described how one question she would ask in class about the logistic population growth model would be hard for students. Given a table of data, students had to calculate r (maximum per capita rate of increase) and K (carrying capacity). The question “combines math and concepts” and required them to “really understand the equation.” Based on prior experience, Gloria explained,

I don’t know if I’ve looked at this historical data on this one recently, but I just know that they’ll get r pretty easily and then K , many of them will just sort of really grasp for straws as to how would I even think through it.

She anticipated how the question would go in class and expected to prompt students to share their reasoning with each other:

I’ll probably prompt them like “I don’t want you just to get the right answer, I want you to explain to each other the thinking you used to get the right answer,” and that will slowly bubble around room probably and take us a little longer to get to. That would be my guess.

These quotes illustrate the integration of PCK components with knowledge of creating opportunities for generative work. Gloria is aware of the concepts that are most difficult for students in the lesson and uses questions that uncover these difficulties.

She also planned to use explicit prompting to remind students to develop explanations with reasoning. She expected that students need specific structure, in the form of verbal prompts, to take the additional steps to generate reasoning (see prior section).

Some generative instructors had designed new in-class work for the target class based on previous observations of difficulties students had with lesson topics. Greg aimed to engage students in a challenging problem that would reveal to him and to the students where they were struggling to understand the topic. He had designed a new task in which students would transcribe and translate a segment of double-stranded DNA to create a polypeptide. He knew from prior lessons in the semester that about half of his students could transcribe and translate a single-stranded piece of DNA. Though this was only his second time teaching the course, he had recognized a key challenge for students was “the directionality issue with DNA.” To address this, he “messes with it a lot on purpose to get them to understand how important it is.” Recognizing that DNA was integral to topics throughout his course, Greg explained,

I just decided you know what, from day one when we talked about biomolecules I’m just going to make directionality a huge deal to the whole class.

In the task in the target class, “the actual template strand is going to be read from right to left,” and students will “have to realize that or they’ll end up with the wrong answer.” In addition, he had planned the subsequent lesson to provide students with another example of the same task with an added point mutation. He explained that,

With this type of activity, students attack it with a very memorize-y sort of approach. And so, I worry that if we do this once, they will look at it in this specific way and so when the test comes around they’ll try to attack it exactly the same way and it won’t go well.

This quote is another illustration of the integration of PCK and knowledge of creating opportunities for generative work, because Greg plans to help students move beyond conceptual difficulties by engaging them in tasks that reveal and challenge their thinking. An instructor who lacked knowledge of creating opportunities for generative work might opt to repeatedly explain the directionality of DNA to students or to spend time finding an exemplary visual representation rather than creating new problems for students.

Some generative instructors had selected in-class work designed by others to target a difficulty they had seen among their students in prior semesters. Gina’s target lesson focused on electrochemical gradients, which can be represented using mathematical equations. Before teaching the target lesson, she anticipated that “everybody is afraid of math” and that students tended to take an algorithmic approach to dealing with equations rather than working to understand the concept underlying the equation. She planned to use a POGIL (process-oriented guided-inquiry learning) activity that she expected to “take the math away” by using arrows to represent the movement of ions across membranes, so students could “just focus on practicing the ideas behind the equation.” She liked POGIL, because it enables “group work with a lot of structure.”

Gina's predictions about how the POGIL activity would facilitate student thinking were supported by her observations during the target class. For example, she noticed that students who had encountered equations about membrane potential in a physics course "had a harder time with the activity," because they knew "how to plug and chug" to get the correct answer, "but they didn't have that supporting understanding of what are the forces." Gina concluded, "I really ended up liking this activity because it built up the 'why.'" In addition to being an example of the integration of PCK and PK, this instance may represent further knowledge development for Gina, because her ideas about what is challenging for students as they learn about membrane potential and how to help them overcome those challenges were corroborated by her experience of teaching this lesson.

Two instructors who exhibited generative instruction were teaching lessons they had not taught previously, and both lamented the lack of knowledge they had about student understanding of the lesson topics. Ginger had taught the course previously and so was able to design some parts of the lesson based on areas where she expected students to struggle. In the postinstruction interview, Ginger explained,

This is one reason I'm so happy with the way this all went is because I always feel so much more comfortable when I know where they're going to kind of trip up. And I didn't anticipate that ... they would focus so much on the story of the bacteria and the virus.

Gail was adopting a lesson designed by an instructor who had extensive experience teaching the course previously. She expressed concern about her lack of knowledge of student understanding of the lesson topics:

Being my first time through the class, this is a part I'm insecure with because I walked in and I teach them and then I'm like "I don't know" until I actually talk with them and then I'm like, "Okay, that's where you're at."

These quotes are noteworthy, because they were unique to generative instructors. Other instructors who similarly lacked knowledge of student understanding of lesson topics did not express concern that it negatively affected their instruction.

Active instructors did not rely on PCK to the same extent in their lesson design. Some active instructors were generally aware of concepts they expected students to find particularly difficult, but there was scant evidence that they designed in-class work specifically to address those difficulties. For example, Adam taught a lesson on DNA as the basis of inheritance. When asked what particular difficulties he expected students to encounter, he said,

I think when you are talking about DNA structure just generally, the structure, it's very complicated. A lot of the stuff, a lot of the slides that I'm going to show tomorrow honestly have too much information on them. And there's, you know, really when talking about DNA structure, I think there's some very specific ideas that I want to get across but I don't know that we've done a good job of being able to get just those ideas across while limiting some of that detail than can cause some confusion among students.

Adam seems to lack detailed knowledge about specifically what aspects of DNA structure students struggle to understand, as well as what instructional strategies could be effective in the face of these difficulties. Furthermore, he does not show evidence of integrating PK to plan a lesson that engages students in generative work around their difficulties.

Active instructors did not exhibit as much knowledge of student understanding about the topics they taught. This difference cannot be explained by differences in teaching experience, because both active and generative instructors had a range of experience (Supplemental Table S1). They also did not gain access to student thinking during the target lesson to the same degree as generative instructors. This was evident when Amelia expressed her concerns that a visual representation she had selected to replace a confusing diagram actually introduced new confusions. These concerns were based solely on her own thinking about the representation, rather than access to how students perceived and thought about the representation, or what was particularly challenging for students about learning the topics illustrated in the representation.

Generative Instructors Used PK While Teaching, Creating Opportunities to Develop New PCK

Generative instructors drew on both components of PK to plan and implement lessons. They not only engaged students in generative work, but also moved around the classroom while students worked and spent the majority of this time talking with small groups of students. Together, these actions provided considerable access to student reasoning about the lesson topics. The teaching practices resulting from these applications of PK created key chances for the instructors to develop PCK. Instructors had opportunities to better understand students' thinking and how student thinking was impacted by instructional practices. Thus, PK facilitated their development of new PCK.

As generative instructors interacted with students, they spent time working to figure out what students were thinking, rather than immediately offering their own explanations about content. This required the instructor to reason in real time about students' thinking, which was often murky and sometimes unanticipated. Greg exemplified these behaviors. He narrated his interactions with a few small groups by saying that he had spent "a lot of that [time] just trying to figure out what they were doing" and was "thinking on the fly" when he encountered answers that he was not expecting.

Greg seemed to build knowledge of student understanding as a result of reasoning about student thinking in interactions with small groups. He appeared to develop some knowledge while engaging in the act of teaching and other knowledge while reflecting after teaching. Some of his PCK was reinforced by his interactions with students. For example, he anticipated a priori that students struggled to understand the directionality of DNA, and particularly what 5' and 3' meant. This knowledge was confirmed in several ways while teaching the target lesson, including when he observed students labeling a polypeptide with 5' and 3' ends after transcribing and translating from a DNA strand. He explained, "you wouldn't label a polypeptide like that if you truly understood what the 5' and 3' meant." Greg considered how he could refine the work he

asked students to do to provide more insight about student thinking,

Anytime you build into the activity opportunities to go wrong, pathways where you can pick the wrong path are diagnostic tools that you can use to figure out what students are thinking and where they're going wrong. So the more opportunities you build into the activity to do that, gives you more data points.

Greg drew on knowledge of monitoring and responding to student thinking, as well as PCK, when he designed questions for students that would reveal to him where they were confused. He then used these "data" about student thinking to improve his own understanding of specific difficulties students encountered (i.e., PCK).

Generative instructors emphasized both the short-term and long-term benefits of hearing students' explanations of their thinking. Gloria described an interaction with a student who was calculating size in a deer population experiencing exponential growth. The student had calculated the population size in year 1 accurately, but had incorrectly calculated the population size in years 2 and 3. This dialogue between Gloria and the interviewer reveals two distinct motivations that Gloria had when she asked the student to "explain to me what you're doing":

Gloria: I had a feeling what she had done wrong, but without doing math very quickly in my head, it was a lot easier for me and better learning experience for her to just explain it to me.

Interviewer: Why do you think her explaining her thinking is a better learning experience?

Gloria: Because she has to think through what she did, and then I can see where she went wrong and hopefully be able to draw out from her, how she can get to the right answer without me telling it to her. But I can't do that if I don't know where she is going wrong in her thinking.

Gloria expected both the student and herself to benefit from the student articulating her thinking. It allowed Gloria to respond appropriately to facilitate the student's learning in the moment. Gloria also explained that interacting with students as they worked was crucial to developing her PCK and therefore her abilities as a college instructor:

[Interacting with students] truly is the way that I learned how to teach because I get this feedback that, you know as a researcher, this kind of qualitative feedback is really different than just getting some numbers on a multiple choice, right, because you have no idea how they're really thinking. So, it's all of these moments in time over all the years that help me decide things that my colleagues could never know about our students if they don't do this.

Amanda, who primarily engaged students in active work (Figure 1), represented a negative case for this theme. Like instructors who engaged students in generative work, she displayed knowledge of monitoring and responding to student thinking. Anytime that students were working during the target

lesson, she was circulating the classroom and prompting students to share their reasoning with her. She said,

So I turn off my [microphone] when they're working on the activity. I walk around the room. I ask them if they have any questions for me to clarify. I eavesdrop on their conversations. I look at their papers. I ask them to explain why they're doing what they're doing.

Though she had the chance to hear students' thinking during the target lesson, we saw little evidence that Amanda was building PCK based on her experiences in the target lesson. She was unable to observe difficulties students experienced while learning the lesson topics, because the questions that she posed to students did not seem to challenge their understanding. This case suggests that knowledge of monitoring and responding to student thinking and engaging students in sufficiently challenging generative work may both be necessary for building PCK from teaching experiences. Amanda seems well equipped to build PCK once she has access to student thinking about challenging questions.

Most active instructors created few opportunities to access student thinking about content during the target lesson and thereby missed opportunities to develop PCK. In lessons in which students mostly engaged in LOCs, instructors interacting with groups heard students' basic understandings, but not their deeper reasoning. The student errors they encountered tended to be errors in recall or basic comprehension, rather than conceptual understanding, so they had few opportunities to begin to understand the conceptual challenges posed by particular content. In other lessons, instructors rarely interacted with students while they worked and therefore missed opportunities to access student thinking.

Additionally, some active instructors quickly answered students' questions without first hearing the students' reasoning. This might help instructors learn *what* students find difficult, but does not help them understand *why* particular content is difficult for students. For example, Alan described a student who asked a question about how codominance and mosaicism were related to each other as "off track." He did not ask any follow-up questions of the student, and his reasoning about the origin of her question was limited to the idea that both codominance and mosaicism were exceptions to traditional Mendelian genetics. Additional questioning may have revealed that the student was making reasonable connections between two genetic mechanisms that can create phenotypes with a speckled or spotted appearance.

Lack of access to student thinking may have resulted in instructors posing questions that were challenging in ways they did not intend. For example, Al learned during the target lesson that students lacked some key background information to answer a question he had been asking for multiple semesters. He learned that students might not know what happens when a seed is germinated in the dark. This information was important to being able to answer a question that was designed to challenge misconceptions about cellular respiration in plants. Al explained,

It just kind of occurred to me because I have asked this question the previous semesters many times and that really hadn't

occurred to me. I'd always thought that they would know what happens when you germinate seeds in the dark, but certainly after talking to that student and she said she didn't remember what happens because the last time they did it was in third grade, but it just occurred to me that maybe a lot of the students in the class just don't know what happens if you put seeds with water in the dark.

Al developed this knowledge of student understanding by calling on a student and listening to her answer. He reflected on this moment in the postinstruction interview as a "revelation to me as an instructor" and explained that

It's very hard for us to put ourselves in the students' shoes or students' frame of mind and know what their prior experience is because the stuff we think are [*sic*] so obvious, for many of our students, are [*sic*] not and it's just a constant reminder that we need to check our assumptions.

This quote demonstrates that Al sees value in understanding students' thinking. Yet he rarely talked to students while they were working during the target lesson. Additionally, the questions he posed tended to be lower-order, multiple-choice questions, so the data he collected did not provide much insight into student reasoning. This instance illustrates how failing to apply knowledge of monitoring and responding to student thinking can result in missed opportunities to develop new PCK.

DISCUSSION

Instructors who engaged students in generative active-learning instruction drew on both topic-specific knowledge about student learning (i.e., PCK) and more generalizable knowledge about how people learn (i.e., PK). These findings build upon an earlier expert–novice study in which expert active-learning instructors were more likely than novices to demonstrate this knowledge as they analyzed videos of other instructors (Auerbach *et al.*, 2018). We examined knowledge in the context of participants' own teaching, allowing an in-depth look at the relationship between teaching knowledge and active-learning instruction, as well as interactions between PK and PCK. Critically, PK may facilitate the development of PCK. Therefore, lack of PK may hinder not only teaching effectiveness, but also the development of new PCK that is also important to effective teaching. This exploratory study suggests hypotheses about teaching knowledge for effective active-learning instruction. These hypotheses will need to be tested in diverse samples of college STEM instructors and across teaching contexts. Specifically, we hypothesize that

1. Knowledge of creating opportunities for generative work, which includes fundamental ideas about how students learn and how learning is facilitated, is necessary for planning and implementing active-learning instruction that leads to high conceptual learning gains.
2. Knowledge of monitoring and responding to student thinking, which places student thinking at the forefront of instruction, is necessary for planning and implementing active-learning instruction that leads to high conceptual learning gains.
3. Instructors who exhibit well-developed PK more efficiently build PCK over time.

4. PCK is necessary for planning and implementing active-learning instruction that leads to high conceptual learning gains and is not necessarily developed through practice alone.

Hypothesis 1 was born from the observation that fundamental principles about how students learn and the application of those principles in real time while teaching were important to the instructional practices of our participants. Participants' ideas about how students learn seemed to influence their perception of the purpose of active-learning instruction. Those instructors who engaged students in multiple opportunities to generate their own explanations and reasoning did so because they thought that students learn by engaging in that type of cognitive work. In contrast, participants exhibiting primarily active instruction used activities to break up lecture and to direct students' attention. The thinking exhibited by these instructors implied that students listening attentively to lectures was key to student learning. This difference in knowledge about how students learn might not be problematic if the instructors had different goals for student learning. For example, instructors who aim to help students memorize information may be well served by composing clear lectures. However, the participants in this study shared a desire to promote deep learning among their students.

Participants who exhibited generative instruction not only asked challenging questions that required students to generate reasoning, they also deliberately prompted students to focus on generating reasoning during class. The idea that instructor prompting is important for effectively implementing generative active-learning instruction is supported by studies of students. Prior work indicates that instructor verbal prompts during class have a real impact on the cognitive work that students do. Knight and colleagues (2013) found that undergraduates who were explicitly prompted by instructors to use reasoning were more likely to have discussions containing evidence-based reasoning than students not prompted in this way. A study of middle school teachers and students similarly indicated that students were more likely to generate their own explanations and reasoning when the teacher gave a question or prompt that asked students to explain or generate new ideas or reasoning (i.e., generative prompt), compared with when the instructor asked questions probing for recall (i.e., active prompt; Chase *et al.*, 2019). Additionally, greater use of generative prompts by teachers were associated with better student performance on questions testing their ability to transfer knowledge to novel problems (Chase *et al.*, 2019). Future work must investigate whether helping undergraduate biology instructors develop knowledge of how people learn and how to apply that knowledge while teaching improves their students' learning.

Hypotheses 2 and 3 recognize that student thinking was central to generative instruction in our findings. Participants exhibiting generative instruction found student thinking interesting and essential to their work. They worked to access student thinking throughout a lesson by talking with multiple students and viewing and interpreting student work. They aimed to ask students challenging questions and felt handicapped in their work as teachers when they lacked insight about what students would find difficult about a topic or task.

Together Hypotheses 2 and 3 encompass both a direct and indirect effect on effective instruction of knowledge of

monitoring and responding to student thinking. We propose that PK directly influences teaching by helping instructors learn how students are progressing with in-class work and giving instructors a chance to provide additional direction, prompting, or support as needed. This aligns well with a model of the process of formative assessment. In this model, a formative assessment prompt (i.e., a question) produces evidence of student understanding that facilitates the diagnosis of in-progress learning, which can then influence instructors to revise instruction and generate feedback (Offerdahl *et al.*, 2018). Hypothesis 2 posits that PK is necessary to carry out the process of formative assessment.

We also propose an indirect effect of PK on effective instruction, in which PK supports the development of PCK, and PCK improves the implementation of active learning. The importance of PK in developing PCK is supported by results from comparative case studies that revealed that an instructor who had more knowledge about how to elicit student thinking during class developed more PCK from the experience of teaching a topic for the first time than did an instructor who lacked this PK (Chan and Yung, 2018). Additionally, teaching professional development programs for K–12 teachers that facilitate teachers' abilities to pay attention to and capitalize on student thinking while teaching have been successful at promoting student-centered instruction, even in first-year teachers (e.g., van Es and Sherin, 2008; Zhang *et al.*, 2011; Santagata and Yeh, 2014). A simple prediction that follows from Hypotheses 2 and 3 is that increasing an instructor's curiosity about, access to, and value for student reasoning will increase student learning gains resulting from active-learning instruction.

Hypothesis 4 is suggested by our work and also by extensive research among K–12 instructors demonstrating associations between PCK, instructional practice, and student learning (e.g., Hill *et al.*, 2005; Kanter and Konstantopoulos, 2010; Sadler *et al.*, 2013; Santagata and Yeh, 2014; Blömeke *et al.*, 2015; Keller *et al.*, 2017). Prior work examining PCK in science education is reviewed in Chan and Hume (2019). This review calls upon researchers to pay particular attention to how PCK is measured. A strength of our study is that we examined PCK across the tasks of teaching in an instructor's own teaching context. There are other valuable approaches to studying PCK (see Chan and Hume, 2019), and we encourage biology education researchers to draw heavily on the plethora of research in K–12 settings to recognize the strengths and shortcomings of these approaches. One interesting prediction that follows from Hypothesis 4 is that teaching professional development that focuses on building PCK will be more effective at increasing student learning gains than will teaching professional development that is not topic specific.

Our research investigates teacher knowledge, which can be considered distinct from teacher beliefs. Teacher beliefs have been referred to in diverse ways, including as orientations, approaches to teaching, and conceptions of teaching (e.g., Kember and Gow, 1994; Trigwell *et al.*, 1994; Devlin, 2006; Friedrichsen *et al.*, 2011). Beliefs are general ideas about teaching that are often elicited by asking questions like "How do you describe your role as a teacher?" (Luft and Roehrig, 2007). Researchers disagree about the relationship between teacher beliefs and instructional practices, with some arguing that beliefs predict practices (e.g., Kember, 1997; Ho *et al.*, 2001),

some arguing that changing practices helps change beliefs (e.g., Guskey, 1986), and others disputing the importance of beliefs in the actual teaching decisions of instructors (e.g., Devlin, 2006; Eley, 2006; Hora, 2014). Current theory in science education views teaching beliefs as "filters" or "amplifiers" between teaching knowledge and teaching practice (Gess-Newsome, 2015; Carlson and Daehler, 2019). Beliefs are considered distinct from teaching knowledge, but are recognized as potentially impacting how knowledge is built and used. In this study, we aimed to elicit specific teacher knowledge by asking questions about actual teaching situations. At times, thinking that is elicited by general questions and thinking elicited by questions about specific contexts may be similar. In those cases, we advocate referring to this thinking as "knowledge," because we want to promote a view of college instructors as learners. We thus find it useful to describe their thinking in the same terms we might use to describe student thinking.

In our study, instructors who reasonably self-identified as active-learning instructors and were seen as such by their colleagues varied considerably in their instructional practices. These observations highlight the need for distinctions beyond active-learning instruction versus traditional instruction. Closely examining participants' instructional practices brought to light important variation that may contribute to the variation we see in learning gains associated with active learning (e.g., Andrews *et al.*, 2011). Some instructors primarily engaged students in questions that required only recall or algorithmic problem solving. Others provided vague questions that likely left students unsure of what they were supposed to do, such as "Talk to your neighbor about this figure." Though students in these courses worked during class, class time may not have been well spent, because the questions did not challenge students to generate their own understandings. Our approach, and other research, indicates that the ICAP framework can be useful for making finer distinctions in instructional practice (Chi and Wylie, 2014; Wiggins *et al.*, 2017; Chase *et al.*, 2019). Finer distinctions in practices, such as contrasting active versus generative active-learning instruction, will be critical to investigating in more detail how teaching practices influence student success.

We also observed that, in many of the classes that included only active work, students opted not to work in small groups when directed to do so. Without a broader look at how collaborative group work was set up early in the semester, it was not always clear why students made this choice. There is some evidence that students learn even more from collaboratively engaging in generative work than when they do this work alone (e.g., Chi *et al.*, 2017). Therefore, supporting instructors in designing and implementing successful group work may be important to student outcomes. We recommend a guide from *CBE—Life Sciences Education* that describes the vast evidence base relevant to group work in undergraduate STEM, including setting up groups, norms and structures for group work, accountability, and task structure for group work (Wilson *et al.*, 2018).

We caution readers about generalizing our results beyond this study and invite increased attention to research on teacher knowledge in undergraduate STEM education. We contrasted instructors in this study based on their practices in a single class session. Analyzing more of their practices may have revealed

important variation (e.g., Stains *et al.*, 2018). We can conclude that participants whose instruction was categorized as “generative” are capable of designing and implementing a lesson that includes multiple opportunities for generative work, but not that they do this for every class session. We cannot conclude that those participants whose instruction was categorized as “active” are incapable of teaching a class with multiple opportunities for generative work. However, instructors often volunteered information in the preinstruction interview about how the target class was similar or different from their other class sessions. Generally, the active instructors described the target lesson as representative of their “typical” class session. Future work would benefit from capturing more than one class session and aiming for class sessions the participants had taught previously, as this would help to minimize variation based on sampling a small number of class sessions.

We also urge caution about drawing conclusions from any single approach to characterizing instructional practices. What occurs between instructors and students in the classroom is complex and variable over time. No single approach can fully capture this complexity. Multiple approaches will be necessary to more fully understand relationships between teacher knowledge, instructional practices, and student outcomes. We triangulated two approaches to analyzing instructional practices, but more could be learned from using additional approaches, such as analyzing the instructor talk exhibited (Seidel *et al.*, 2015), the 21 specific practices documented in the PORTAAL (Eddy *et al.*, 2015), or the participation patterns in classroom discourse provided by Equity Quantified in Participation (Reinholz and Shah, 2018).

Many of our participants had some experience with education research, and we also observed that some participants referred to specific education research papers that had informed their thinking and practice in important ways. This could suggest that repeated exposure to education research helps instructors gain PK and PCK. On the other hand, not all participants exhibiting generative instruction had education research experience, and all but one of the instructors exhibiting active instruction reported this background. Thus, experience with education research, in the form of education research training, presenting at a conference, or publishing a paper, does not necessarily result in effective planning and implementation of active-learning instruction. Nonetheless, this exposure may have positive impacts. Future research can better untangle the role of education research literature and experience in supporting the development of important teaching knowledge.

CONCLUSION

As a biology education community, we continue to call for college biology instructors to adopt active learning, both at the national level and in many local efforts within institutions and departments. In our zeal to move faculty away from traditional lecture methods, we may have inadvertently communicated that anyone can use active-learning instruction and achieve impressive outcomes for their students. This glosses over the reality that implementing these practices effectively may depend on a wealth of knowledge beyond what instructors use to compose and deliver lectures, and beyond the knowledge they use as practicing biologists. Furthermore, it may be common for instructors to literally interpret “active learning” as

engaging students in activity, rather than creating opportunities for generative work. This paper takes a small step toward unpacking the complex knowledge active-learning instructors may need. It also models one way to more specifically define active-learning instruction (i.e., the ICAP framework). We hope that future research takes this agenda further, capitalizes on work in other STEM disciplines and in K–12 settings, and intersects with work to establish critical components of effective active-learning instruction (e.g., Eddy *et al.*, 2015; Stains and Vickrey, 2017; Offerdahl *et al.*, 2018).

ACKNOWLEDGMENTS

Thank you to our research participants for generously offering their time and insights. Thanks to Michelle A. Ziadie, who helped with “Blooming” work students completed during class, and to Alex Waugh and Ian John for research assistance. Thanks also to the University of Georgia BERG (Biology Education Research Group) for ongoing feedback and support, and to Julie Stanton and Jennifer Thompson for seeing us through the darker days of coding complex data. We also thank two anonymous reviewers and the monitoring editor for feedback that improved the quality of this article. Partial support for this work was provided by a seed grant from the University of Georgia’s Owens Institute for Behavioral Research and by the National Science Foundation’s Improving Undergraduate Education (IUSE) program under award no. 1504904. This material is based on work conducted by E.F.G., who was supported in part by the National Science Foundation under REU grant number 1659423. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Owens Institute for Behavioral Research.

REFERENCES

- Abraham, J. K., Perez, K. E., & Price, R. M. (2014). The Dominance Concept Inventory: A tool for assessing undergraduate student alternative conceptions about dominance in Mendelian and population genetics. *CBE—Life Sciences Education*, 13(2), 349–358.
- Alonzo, A. C., & Kim, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching*, 53(8), 1259–1286.
- Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE—Life Sciences Education*, 10(4), 394–405.
- Anfara, V. A., Jr., Brown, K. M., & Mangione, T. L. (2002). Qualitative analysis on stage: Making the research process more public. *Educational Researcher*, 31(7), 28–38.
- Arneson, J. B., & Offerdahl, E. G. (2018). Visual literacy in Bloom: Using Bloom’s taxonomy to support visual learning skills. *CBE—Life Sciences Education*, 17(1), ar7.
- Auerbach, A. J., Higgins, M., Brickman, P., & Andrews, T. C. (2018). Teacher knowledge for active-learning instruction: expert–novice comparison reveals differences. *CBE—Life Sciences Education*, 17(1), ar12.
- Auerbach, A. J. J., & Andrews, T. C. (2018). Pedagogical knowledge for active-learning instruction in large undergraduate biology courses: A large-scale qualitative investigation of instructor thinking. *International Journal of STEM Education*, 5(1), 19.
- Ball, D.L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407.
- Birks, M., & Mills, J. (2011). *Grounded theory: A practical guide*. London: Sage.

- Blömeke, S., Hoth, J., Döhrmann, M., Busse, A., Kaiser, G., & König, J. (2015). Teacher change during induction: Development of beginning primary teachers' knowledge, beliefs, and performance. *International Journal of Science and Mathematics Education, 13*, 287–308.
- Carlson, J., & Daehler, K. R. (2019). The refined consensus model of pedagogical content knowledge in science education. In Hume, A., Cooper, R., & Borowski, A. (Eds.), *Repositioning PCK in teachers' professional knowledge for teaching science* (pp. 77–92). Sydney, Australia: Springer.
- Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In Hume, A., Cooper, R., & Borowski, A. (Eds.), *Repositioning PCK in teachers' professional knowledge for teaching science* (pp. 3–76). Sydney, Australia: Springer.
- Chan, K. K. H., & Yung, B. H. W. (2015). On-site pedagogical content knowledge development. *International Journal of Science Education, 37*(8), 1246–1278.
- Chan, K. K. H., & Yung, B. H. W. (2018). Developing pedagogical content knowledge for teaching a new topic: More than teaching experience and subject matter knowledge. *Research in Science Education, 48*(2), 233–265.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Thousand Oaks, CA: Sage.
- Chase, C. C., Marks, J., Malkiewich, L. J., & Connolly, H. (2019). How teacher talk guidance during invention activities shapes students' cognitive engagement and transfer. *International Journal of STEM Education, 6*(1), 14.
- Chi, M. T., Kang, S., & Yaghmourian, D. L. (2017). Why students learn more from dialogue- than monologue-videos: Analyses of peer interactions. *Journal of the Learning Sciences, 26*(1), 10–50.
- Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243.
- Cooper, M. M. (2016). It is time to say what we mean. *Journal of Chemical Education, 93*(5), 799–800.
- Cresswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice, 39*(3), 124–130.
- Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE—Life Sciences Education, 7*(4), 368–381.
- Dancy, M., Henderson, C., & Turpen, C. (2016). How faculty learn about and implement research-based instructional strategies: The case of peer instruction. *Physical Review Physics Education Research, 12*(1), 010110.
- Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education, 34*, 12–25.
- Devlin, M. (2006). Challenging accepted wisdom about the place of conceptions of teaching in university teaching improvement. *International Journal of Teaching and Learning in Higher Education, 18*(2), 112–119.
- 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.
- Eley, M. G. (2006). Teachers' conceptions of teaching, and the making of specific decisions in planning to teach. *Higher Education, 51*(2), 191–214.
- 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 USA, 111*(23), 8410–8415.
- Friedrichsen, P., Van Driel, J. H., & Abell, S. K. (2011). Taking a closer look at science teaching orientations. *Science Education, 95*(2), 358–376.
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. In Berry, A., Friedrichsen, P., & Loughran, J. (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). New York: Routledge.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick, NJ: AldineTransaction.
- Grossman, P. L., & Richert, A. E. (1988). Unacknowledged knowledge growth: A re-examination of the effects of teacher education. *Teaching and Teacher Education, 4*(1), 53–62.
- Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal, 29*(2), 75–91.
- Guskey, T. R. (1986). Staff development and the process of teacher change. *Educational Researcher, 15*(5), 5–12.
- Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science, 332*(6034), 1213–1216.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.
- Ho, A., Watkins, D., & Kelly, M. (2001). The conceptual change approach to improving teaching and learning: An evaluation of a Hong Kong staff development programme. *Higher Education, 42*(2), 143–169.
- Hora, M. T. (2014). Exploring faculty beliefs about student learning and their role in instructional decision-making. *Review of Higher Education, 38*(1), 37–70.
- Johnson, E. M., & Larsen, S. P. (2012). Teacher listening: The role of knowledge of content and students. *Journal of Mathematical Behavior, 31*(1), 117–129.
- Kanter, D. E., & Konstantopoulos, S. (2010). The impact of a project-based science curriculum on minority student achievement, attitudes, and careers: The effects of teacher content and pedagogical content knowledge and inquiry-based practices. *Science Education, 94*(5), 855–887.
- Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching, 54*(5), 586–614.
- Kember, D. (1997). A reconceptualisation of the research into university academics' conceptions of teaching. *Learning and Instruction, 7*(3), 255–275.
- Kember, D., & Gow, L. (1994). Orientations to teaching and their effect on the quality of student learning. *Journal of Higher Education, 65*(1), 58–74.
- Knight, J. K., Wise, S. B., & Southard, K. M. (2013). Understanding clicker discussions: Student reasoning and the impact of instructional cues. *CBE—Life Sciences Education, 12*(4), 645–654.
- König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations. A video-based assessment approach. *Teaching and Teacher Education, 38*, 76–88.
- Luft, J. A., & Roehrig, G. H. (2007). Capturing science teachers' epistemological beliefs: The development of the teacher beliefs interview. *Electronic Journal of Science Education, 11*(2).
- Magnusson, S., Krajcik, J., & Borke, H. (1999). Nature, sources, and development of pedagogical content knowledge. In Gess-Newsome, J., & Lederman, N. G. (Eds.), *Examining pedagogical content knowledge* (pp. 95–132). New York: Kluwer.
- Maher, J. M., Markey, J. C., & Ebert-May, D. (2013). The other half of the story: Effect size analysis in quantitative research. *CBE—Life Sciences Education, 12*(3), 345–351.
- Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education, 102*(3), 346–374.
- Morine-Dershimer, G., & Kent, T. (1999). The complex nature and sources of teachers' pedagogical knowledge. In Gess-Newsome, J., & Lederman, N. G. (Eds.), *Examining pedagogical content knowledge* (pp. 21–50). New York: Kluwer.
- Nelson, K. P., & Edwards, D. (2015). Measures of agreement between many raters for ordinal classifications. *Statistics in Medicine, 34*(23), 3116–3132.
- Novick, L. R., & Catley, K. M. (2013). Reasoning about evolution's grand patterns: College students' understanding of the tree of life. *American Educational Research Journal, 50*(1), 138–177.
- Offerdahl, E. G., McConnell, M., & Boyer, J. (2018). Can I have your recipe? Using a fidelity of implementation (FOI) framework to identify the key ingredients of formative assessment for learning. *CBE—Life Sciences Education, 17*(4), es16.

- Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38*(3), 261–284.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.
- Petrou, M., & Goulding, M. (2011). Conceptualising teachers' mathematical knowledge in teaching. In Rowland, T., & Ruthven, K. (Eds.), *Mathematical knowledge in teaching* (pp. 9–25). Dordrecht, Netherlands: Springer.
- Reinholz, D., & Shah, N. (2018). Equity analytics: A methodological approach for quantifying participation patterns in mathematics classroom discourse. *Journal for Research in Mathematics Education, 49*(2), 140–177.
- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal, 50*(5), 1020–1049.
- Saldaña, J. (2013). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage.
- Santagata, R., & Yeh, C. (2014). Learning to teach mathematics and to analyze teaching effectiveness: Evidence from a video- and practice-based approach. *Journal of Mathematics Teacher Education, 17*(6), 491–514.
- Seidel, S. B., Reggi, A. L., Schinske, J. N., Burrus, L. W., & Tanner, K. D. (2015). Beyond the biology: A systematic investigation of noncontent instructor talk in an introductory biology course. *CBE—Life Sciences Education, 14*(4), ar43.
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22*(2), 63–75.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–23.
- Smith, M. K., Jones, F. H., 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.
- Smith, P. S., Esch, P. R., Hayes, M. L., & Plumley, C. L. (2016). Developing and testing a method for collecting and synthesizing pedagogical content knowledge. Paper presented at: NARST Annual International Meeting (Baltimore, MD).
- Speer, N. M., & Wagner, J. F. (2009). Knowledge needed by a teacher to provide analytic scaffolding during undergraduate mathematics classroom discussions. *Journal for Research in Mathematics Education, 53*(3), 530–562.
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21–51.
- Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., ... Levis-Fitzgerald, M. (2018). Anatomy of STEM teaching in North American universities. *Science, 359*(6383), 1468–1470.
- Stains, M., & Vickrey, T. (2017). Fidelity of implementation: An overlooked yet critical construct to establish effectiveness of evidence-based instructional practices. *CBE—Life Sciences Education, 16*(1), rm1.
- Trigwell, K., Prosser, M., & Taylor, P. (1994). Qualitative differences in approaches to teaching first year university science. *Higher Education, 27*(1), 75–84.
- van Driel, J. H., Berry, A., & Meirink, J. (2014). Research on science teacher knowledge. In Lederman, N.G., & Abell, S. K. (Eds.), *Handbook of research on science education* (Vol. 2, pp. 862–884). New York: Routledge.
- van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education, 24*(2), 244–276.
- Wagner, J. F., Speer, N. M., & Rossa, B. (2007). Beyond mathematical content knowledge: A mathematician's knowledge needed for teaching an inquiry-oriented differential equations course. *Journal of Mathematical Behavior, 26*(3), 247–266.
- Wiggins, B. L., Eddy, S. L., Grunspan, D. Z., & Crowe, A. J. (2017). The ICAP active learning framework predicts experimentally assessed learning gains for intensely active classroom experiences. *AERA Open, 3*, 2332858417708567.
- Wilson, K. J., Brickman, P., & Brame, C. J. (2018). Group work. *CBE—Life Sciences Education, 17*(1), fe1.
- Zhang, M., Lundeborg, M., Koehler, M. J., & Eberhardt, J. (2011). Understanding affordances and challenges of three types of video for teacher professional development. *Teaching and Teacher Education, 27*(2), 454–462.
- Ziadie, M. A., & Andrews, T. C. (2018). Moving evolution education forward: A systematic analysis of literature to identify gaps in collective knowledge for teaching. *CBE—Life Sciences Education, 17*(1), ar11.