

Meeting students halfway: Increasing self-efficacy and promoting knowledge change in astronomy

Janelle M. Bailey,^{1*} Doug Lombardi,¹ Jacqueline R. Cordova,² and Gale M. Sinatra³

¹*Department of Teaching and Learning, Temple University,*

1301 Cecil B. Moore Ave., Philadelphia, Pennsylvania 19122, USA

²*Department of Educational Psychology and Higher Education, University of Nevada Las Vegas,*
4505 S. Maryland Parkway, MS 3003, Las Vegas, Nevada 89145, USA

³*Rossier School of Education, University of Southern California, Waite Phillips Hall,*
3470 Trousdale Parkway, Los Angeles, California 90089, USA

(Received 23 May 2016; revised manuscript received 14 October 2017; published 26 December 2017)

Two motivational factors—self-efficacy and interest—may be especially relevant to deepening students’ understanding of astronomy. We examined the relationship between students’ self-efficacy for, interest in learning about, and changes in their knowledge of stars, as measured by the Star Properties Concept Inventory (SPCI). Approximately 700 undergraduate students taking introductory astronomy responded to surveys at the start and end of their semester-long course. A sequential multiple regression analysis showed that self-efficacy post explains an appreciable percentage of variance in SPCI posttest scores, more than twice the percentage explained by all the pretest variables (SPCI, self-efficacy, and interest) combined. Knowledge and self-efficacy improved significantly over instruction; interest did not. Follow-up analyses revealed that instructors whose classes increased in self-efficacy also had the greatest increases in knowledge scores. Interviews with these instructors suggest they provide their students with more opportunities for mastery experiences with elaborated, performance-related feedback, as well as strong positive verbal persuasion and vicarious experiences through peer instruction. Through increased understanding of the relationship between motivational constructs (e.g., self-efficacy, interest) and knowledge, we can both improve our models and better inform instruction.

DOI: [10.1103/PhysRevPhysEducRes.13.020140](https://doi.org/10.1103/PhysRevPhysEducRes.13.020140)

I. INTRODUCTION

As college science courses are reconceptualized to better facilitate student understanding, researchers and instructors are paying increased attention to the motivational factors that impact both (a) the initial acquisition and construction of knowledge and (b) the reconstruction of knowledge from naïve to more sophisticated ideas (often referred to as conceptual change). Research on knowledge change, in particular, began by assuming a primarily rational, cognitive approach, but has grown to include a variety of motivational constructs [1–3]. One of these constructs, *self-efficacy*, is a person’s perception of his or her capability to carry out the actions required to accomplish a given task [4]. Self-efficacy is thought to be domain or even task specific [5,6], thus research on self-efficacy in a variety of domains is warranted. Another motivational construct addressed in some conceptual change models is *interest*. Sometimes equated to a sense of curiosity about a topic,

interest can impact attention, goals, and learning; see Ref. [7] and references therein. Self-efficacy and interest can affect a student’s engagement with a topic and, when high, may facilitate conceptual change [1].

Beginning with the cognitive reconstruction of knowledge model (CRKM) [1], which posits that motivational factors could have a strong influence on knowledge change, we specifically examined self-efficacy and interest as two motivational factors that may have particular relevance in the astronomy classroom. To characterize self-efficacy, we rely on Bandura’s [8] social cognitive theory; for interest, we base the construct on Hidi and Renninger’s [7] model of interest development.

To date, little research has been conducted on self-efficacy or interest, or their impacts on changes in knowledge, within astronomy [9–11]. Undergraduate astronomy provides a unique context in which to investigate these topics: It is taken by many students across the U.S. [12], including a large fraction who may become teachers in the future [13]. Astronomy is often taken by nonscience majors to satisfy broad science requirements [14,15], perhaps more so than some other fields such as physics or chemistry. The broad applicability across university students is an important aspect for science education research, as the implications may hold promise in other disciplines as well. A core concept within astronomy is stars—they are the building

*To whom all correspondence should be addressed.
janelle.bailey@temple.edu

Published by the American Physical Society under the terms of the *Creative Commons Attribution 3.0 License*. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

blocks of stellar systems and of galaxies and are among the most commonly included topics in an astronomy syllabus [16], thus providing a relevant and popular topic for study.

The present study was designed specifically to look at the interaction of self-efficacy, interest, and knowledge about the topic of stars. Although we have used knowledge reconstruction (i.e., conceptual change) as a theoretical basis for the selection of the variables self-efficacy and interest, these factors are important in learning situations involving simple construction of knowledge, and also a combination of knowledge construction and reconstruction. We therefore begin with a review of the relevant conceptual change literature that informs our work, before continuing with self-efficacy and interest within astronomy education. We next describe the methods of the present study, followed by the results, discussion, limitations, and conclusions (including future research and implications).

II. BACKGROUND

In this section we will identify and discuss the background research relevant to our study. First we discuss the three pertinent constructs (conceptual change, self-efficacy, and interest) by exploring the terms and then discussing each in the context of astronomy education research. We next address the use of concept inventories as a measure of knowledge change in astronomy. This is followed by a discussion of instructional strategies in astronomy that have been designed or used to elicit knowledge change.

A. Conceptual change

Educators have long been concerned with students' naïve conceptions about scientific content, and more specifically, how we can help students move from naïve ideas to a more robust scientific understanding. Conceptual change is often involved in this transition of understanding. Conceptual change researchers investigate, among other things, various change processes, the factors that influence the degree to which change occurs, and the instructional methods and environments that facilitate change [17].

1. Cognitive reconstruction of knowledge model (CRKM)

Conceptual change models initially incorporated only rational, cognitive processes [18]; such models were later referred to as “cold conceptual change.” These models stated or implied the change process involved active engagement by the learner. Specifically, the learner would explicitly compare existing understanding to a new idea and choose between them. Criticisms of such models, and subsequent revisions, demonstrated that conceptual change may not occur in an explicit manner [19–21]. Pintrich *et al.* [2] described how motivational and affective constructs could affect a student's ability to undergo conceptual change. This started a “warming trend” [3] because

researchers consider motivation and affect to be extrarational (i.e., “hot”) constructs.

Dole and Sinatra [1], in response to the call for the inclusion of hot constructs, developed the CRKM, combining elements of conceptual change models from cognitive psychology, science education, and social psychology. The CRKM posits that conceptual change is influenced by far more than previous models had included, identifying five major potential factors. The first factor is characteristics of the learner's prior conceptions—what is it that she already knows? The extent to which a prior conception is robust (i.e., how strong it is, how coherent it is, and the learner's commitment to it) will influence the likelihood of conceptual change. The second factor is a learner's motivation to change. The CRKM includes multiple motivational constructs, including the learner's dissatisfaction with their prior conceptions, the social context in which the potential conceptual change may take place, the learner's need for cognition, and the personal relevance of the issue. Dole and Sinatra [1] consider self-efficacy and interest to be critical components of personal relevance.

The third major factor relates to the characteristics of the message itself, i.e., the new conception that is to be learned. A new conception is unlikely to take hold if it is not comprehensible to the learner; likewise, it should be coherent and plausible. Additionally, the message should be rhetorically compelling (“convincing and persuasive”; [1], p. 120). These learner and message characteristics interact with one another but are filtered through the fourth major factor: the level of cognitive engagement with which the learner interacts with the message. Finally, a peripheral cue (the fifth major factor) may further influence the learner's cognitive engagement, in either a positive or negative manner. Peripheral cues, as described in prior research and maintained in the CRKM, might include “an attractive, credible, or trustworthy source; ...a pleasant context... a simple message that [students] understand easily” [1], (p. 117).

The factors described by Dole and Sinatra [1] interact in a dynamic manner; unlike some other models, the conceptual change process is understood here to be nonlinear. Dole and Sinatra also acknowledge that the motivation factor, in particular, may be broader than what they describe. Pintrich *et al.* [2], in their description of hot conceptual change, indicate they “believe that, in terms of individual conceptual change in the classroom, ... the actual content of students' theories and models is influenced by personal, motivational, social, and historical factors” (p. 170), and include self-efficacy and personal interest as two of these factors. Before we turn to a discussion of self-efficacy and interest, we will look at the treatment of conceptual change within our selected subject of astronomy.

2. Conceptual change in astronomy

Astronomy has been used as a context for conceptual change research for more than 20 years; see, e.g.,

Refs. [21–23] and references therein. Many of these studies have taken place within elementary classrooms with relevant topics, such as the cause of the day-night cycle, apparent motions of celestial objects, and lunar phases. Research on conceptual change in undergraduate astronomy is growing, though it often uses a different approach than the K–12 investigations [10]. For example, research at the undergraduate level has focused on the development and testing of specific assessments, such as concept inventories, to measure knowledge change [24–27] or on using such assessments to compare the results of different instructional strategies [28–32].

Most conceptual change studies in undergraduate astronomy, such as those described above, take the cold conceptual change approach, largely based on the conceptual change model proposed by Posner *et al.* [18], rather than warmer models such as the CRKM [1]. The exploration of different CRKM learner and message characteristics is an active research area in education, but inquiries into motivational constructs (such as self-efficacy and interest) within astronomy are rare. Studies addressing these issues in an introductory general education college course are nearly nonexistent, and to our knowledge, no study has addressed both of these constructs in this context in the same study [10]. Thus, there is a need to explore how the findings of prior research regarding the importance of motivational constructs play out in tertiary astronomy courses.

Existing research in other contexts provides some evidence that instructional interventions designed to foster conceptual change may concurrently impact motivational and affective factors as well as knowledge. For example, Broughton *et al.* [33] found that 5th and 6th grade students demonstrated shifts in both emotion and conceptual understanding about Pluto’s reclassification following a targeted intervention. Likewise, Cordova and colleagues [34,35] found that in addition to conceptual change, undergraduate education majors exhibited changes in self-efficacy and confidence about that knowledge following an intervention about the cause of seasons. Although these studies offer evidence that motivational beliefs may shift after an instructional intervention, we know of no research that has investigated changes in self-efficacy and interest simultaneously within the context of undergraduate astronomy. The next two sections will further describe each of these facets and discuss their presence in astronomy education research.

B. Self-efficacy

1. Defining self-efficacy and its sources

Bandura’s [36] behavioral theory states self-efficacy “is the conviction that one can successfully execute the behavior required to produce” specified effects (p. 193). High self-efficacy can guide a person toward putting effort into certain activities, whereas low self-efficacy can lead

individuals to avoid activities they do not believe they are capable of doing. Subsequently, high self-efficacy may lead to greater cognitive engagement, which Dole and Sinatra [1] suggest is a mediator between the learner, the message, and the degree of conceptual change.

What factors can cause changes (particularly improvements) in self-efficacy? Bandura [4] discusses four main sources of self-efficacy: (a) mastery experiences, (b) vicarious experiences, (c) verbal persuasion, and (d) physiological and affective states. Bandura considered mastery experiences—the actual successful completion of tasks—the most important factor affecting self-efficacy, “because they provide the most authentic evidence of whether one can muster whatever it takes to succeed” [4], (p. 80). Successful performance of a task can increase self-efficacy, whereas performance failure—especially early on—potentially lowers it. Subsequent research, however, has shown that mastery experiences may not be the most influential source of self-efficacy for women as it often is for men (see, e.g., Ref. [37] and references therein); for example, vicarious experiences were more important for women in passing introductory physics [37]. The vicarious experience (sometimes also called modeling), in contrast to mastery, refers to seeing someone else, with whom one can identify, successfully completing the same task. The comparison of one’s own performance with others’ or the prediction of future performance in this way can influence self-efficacy in a manner similar to mastery experiences. The third source, verbal persuasion, refers to encouragement or discouragement from another, with discouragement likely lowering self-efficacy and encouragement having the potential to increase self-efficacy. Positive verbal persuasion can “mobilize greater effort and sustain it” toward an activity or task [4], (p. 101). Finally, physiological and affective states relate to one’s interpretation of the physiological cues they experience (e.g., emotional responses to classroom events, such as anxiety and/or nervousness) and may also impact self-efficacy. Physiological factors tend to have greater influence for tasks that involve physical activity, health functioning, stamina, and the like. Any change in self-efficacy is determined not only by the presence of a source itself (i.e., mastery experience, vicarious experience, verbal persuasion, or a physiological factor) but also by how the individual internalizes and interprets information that arises from that source.

2. Domain-specificity: Self-efficacy in astronomy

Research on self-efficacy’s impact covers a wide variety of topics, associated with numerous academic settings [5]. However, several researchers theorize that self-efficacy is largely domain and even task specific [5,6,38], making measurement of self-efficacy more challenging. Britner and colleagues have looked at levels of self-efficacy for middle school [39] and high school students [40] and what differences might exist between science domains. Other

self-efficacy research focuses on college courses, including domains such as physics [37,41,42], or focuses on the self-efficacy for teaching science rather than learning science [43–46]. Investigating self-efficacy in different domains will help researchers understand the nature of this complex variable, and can lead instructors toward understanding how self-efficacy in their own content areas might be improved.

Self-efficacy for astronomy learning has received little attention, despite recent efforts to increase students' cognitive engagement through the inclusion of learner-centered curricular materials and instructional strategies. In the few cases where self-efficacy has been examined within an astronomy course, researchers often measured self-efficacy for general science learning (or science teaching), rather than specifically for astronomy learning (or astronomy teaching). Carter [47] investigated sources of self-efficacy toward an astronomy laboratory for undergraduate non-science majors through a mixed-methods research design. She found through interviews that the labs appeared to contribute to self-efficacy *in astronomy* primarily through verbal persuasion, but did not affect *general science* self-efficacy as measured by a survey. Straits and Wilke [32] observed a decrease in general science self-efficacy after an activities-based introductory astronomy course, as determined by a subset of survey questions. Slater *et al.* [48] investigated changes in content knowledge, attitudes, and self-efficacy during an astronomy course for preservice elementary teachers, using in part the Science Teaching Efficacy Belief Instrument, Form B (STEBI-B) [49]. Although the preservice teachers' content knowledge improved significantly, their self-efficacy toward (general) science teaching and attitudes toward science did not. All of these studies involved instructional designs intended to facilitate high cognitive engagement, and two [32,47] measured knowledge change through the Astronomy Diagnostic Test, a multiple-choice measure of general astronomy content knowledge [50]. However, none of these researchers made any explicit connections to warm conceptual change using a model such as the CRKM [1].

Cordova and colleagues [34,35,51] have taken the domain and task specificity of self-efficacy into account, by investigating the roles of self-efficacy and confidence in prior knowledge on conceptual change about the cause of seasons with undergraduate education students. In this work, participants responded to questions relating to specific academic tasks (e.g., reading) within the context of learning about seasonal change. The researchers found that participants classified as having medium self-efficacy showed significantly greater pre-to-posttest gains in understanding of the content compared to either the high or low self-efficacy groups [51]. There was also a significant interaction among all three variables (self-efficacy, confidence in prior knowledge, and time) on the number of alternative conceptions present in participants' posttests,

suggesting the relationship among these factors may be complex [35]. This research relating to task-specific self-efficacy for learning about a particular astronomy topic influenced the design of our study and informed the development of our research questions.

C. Interest

1. Defining interest

Interest, like self-efficacy, is a component of motivation that can lead to higher cognitive engagement [7] and, subsequently, facilitate stronger conceptual change [1]. "Interest refers to a learner's predisposition to reengage particular disciplinary content...over time and the psychological state that accompanies this engagement" [52], (p. 106). With respect to the CRKM, interest may directly influence the level at which the learner engages with the incoming message, and so may be an important motivational characteristic. Interest is often measured through self-report instruments using Likert-scale items, where participants rate their level of interest on given topics or tasks, and may be used to predict participants' behaviors [38]. The connection between interest and conceptual change is an active area of research (see, e.g., Refs. [53,54], and references therein). Like self-efficacy, interest is a domain-specific construct, directed toward an object or topic, and so exploration of interest in new domains is warranted.

2. Domain-specificity: Interest in astronomy

To date, the research focus for interest in astronomy has also been at the K–12 level, with both students and teachers. Reis, Atamian, and Renzulli [55] studied 10 students in a gifted and talented program (grades 5–7) who were given binoculars, telescopes, and a self-directed astronomy unit for study at home with the help of their parents. The researchers found that not only did the students experience growth in the skills targeted by the instructional unit, most also demonstrated increased interest in astronomy. Two studies have also investigated the interests of secondary students relating to astronomy. The first involved approximately 3000 15-year olds in Northern Ireland, who completed a Likert-scale interest survey about various subtopics from astronomy, biology, chemistry, and physics, before and after instruction on the "Earth in Space" portion of the National Curriculum [56]. Astronomy subtopics received the highest interest ratings both pre and postinstruction, although the authors did not report significance testing of differences over time or between topics. The second study surveyed 152 Canadian ninth graders about their interests in topics relating to astronomy and space science [57]. Both boys and girls indicated the strongest interest in black hole formation, extraterrestrials or unidentified flying objects (UFOs), and living in space. Whereas these studies help us understand something about

students' interest in astronomy topics, they do not tell us anything about how these interests interact with students' understanding or their self-efficacy for learning about the topics.

In the early 1990s, three graduate theses documented the study of inservice teachers' interest in (and in some cases knowledge about) astronomy. Shore [58] found that a dozen Grade 2 teachers who were teaching a 10-week astronomy unit significantly increased their interest in astronomy compared to a control group of Grades 1–3 teachers not teaching astronomy, even when controlling for prior interest. However, there was no difference between the groups (teaching vs not teaching) in conceptual knowledge scores, and both groups held similar alternative conceptions. Rutkowski [59] found that junior high teachers had interest in astronomy, believed students were interested in it, and felt astronomy was an important part of the curriculum; at the same time, however, teachers were reluctant to increase the amount of curricular time for astronomy. Slater [60] showed that 25 middle and high school teachers enrolled in a constructivist astronomy and space science course experienced significant increases in their attitudes about, value of, and interest in teaching astronomy, as well as growth in their content knowledge. Although all three studies discussed knowledge change, the projects did not examine the specific relations between interest and instruction.

D. Concept inventories as a measure of changes in knowledge

For nearly three decades, concept inventories have provided a way for researchers and instructors to measure the presence of alternative conceptions and how those conceptions change as a result of instruction [61,62]. The development of these instruments typically includes detailed qualitative data collection, such as through interviews and open-ended questions, and analysis prior to the creation and validation of multiple-choice instruments [61,63]. Concept inventories have been used extensively in undergraduate science courses to measure changes in students' understanding, whether through knowledge construction or conceptual change, over some instructional intervention (such as a unit, curriculum, or semester-long course; in physics, for example, see Ref. [64] and references therein).

The key difference between a concept inventory and some other assessments (e.g., a cumulative final exam) is that the concept inventory focuses in on a single or narrow set of concepts, and includes several questions that get at the same idea. For example, the Force Concept Inventory (FCI) [65] includes 29 questions around six areas (kinematics, first law, second law, third law, superposition principle, and kinds of force), all of which relate to Newtonian mechanics. When used as a whole, a concept inventory can provide a picture about a students' likely

understanding of the concept at hand. When used over time, we can get a sense of how that understanding has changed (for just a few examples within astronomy and physics, see Refs. [28,30,31,64,66]).

E. Instruction to elicit knowledge change

Concept inventories have been used as a measurement of knowledge change as a result of instructional interventions. In a meta-analysis of physics education reform efforts, Hake looked at changes in FCI scores across traditional and what he termed “interactive engagement” courses, and found that the latter had normalized gain scores (a ratio of the actual to maximum potential gain) that were, on average, two standard deviations above those of the former [66]. Many subsequent studies have found similar results.

But what constitutes interactive engagement strategies, and how do they facilitate change? There is no single method for teaching for conceptual change, however some commonalities can be found. Cognitive conflict is a key part of several conceptual change strategies, harkening back to Piagetian-based approaches [67]. Leach and Scott [68] describe several perspectives on teaching for conceptual change, whereas Jonassen [69] focuses on different types of modeling and Linn [70] addresses the conflict between having students “extinguish the ideas that experts dispute, or... distinguish among normative and personally constructed views” (p. 694). The *Lecture Tutorials for Introductory Astronomy* [71], which have become quite popular within introductory astronomy courses, were designed from a constructivist approach and include tasks that deliberately elicit prior student knowledge (many of which are known from research to be scientifically inaccurate) and then help students resolve any conflicts [72]. These and other strategies that create opportunities for the construction of robust scientific understanding have become an integral part of the wider reform efforts within science education, particularly at the undergraduate introductory level.

III. THE PRESENT STUDY

The CRKM [1] claims that motivational characteristics—such as self-efficacy and interest—influence the degree to which a learner experiences conceptual change, but neither has been investigated sufficiently. When combined with the suggestion of the domain and task specificity of these constructs [5,6], research on their interactions within a context not previously studied is warranted. The context of astronomy—specifically the topic of stars—is particularly fruitful because of the popularity of the course at the undergraduate level [12], suggesting some initial interest. Another reason for astronomy as a potentially useful context is that anecdotal evidence suggests students may choose astronomy over other science courses, such as chemistry or physics, because they perceive astronomy as being “easier.”

which might contribute to high pretest self-efficacy ratings. But how do self-efficacy and interest impact students' learning of astronomy? To address these issues, we formulated the following research questions:

- (1) How does postinstruction self-efficacy for learning astronomy relate to changes in understanding of star properties, a topic common to many introductory undergraduate astronomy courses?
- (2) (a) How do understanding of star properties, self-efficacy for learning about star properties, and interest for learning about star properties change after semester-long instruction and (b) how are they affected by instructional design and implementation?

We hypothesized that postinstruction self-efficacy (compared to interest) would be more strongly related to change in understanding of star properties over the course of instruction. Based upon prior research in physics and astronomy education [41] and the study's context, we further hypothesized that all three variables (understanding, self-efficacy, and interest) would increase significantly over a semester-long course. We also predicted that an increased attention to mastery and vicarious experiences within the instructional designs and implementation could lead to improvements in these variables over the courses.

IV. METHODS

The present study builds upon previous work [34,35,51] that investigates students' self-efficacy for tasks associated with and interest in learning about seasonal change. Whereas those studies were completed in an experimental design setting in which students participated as a way of satisfying requirements for an educational psychology course, we combined the focus on motivational variables

with other work around content knowledge and conducted the study in the natural setting of an introductory astronomy course. This focus allowed us to investigate how the relationships between the selected motivational variables and knowledge change, as theorized in Dole and Sinatra's CRKM [1], play out in the type of context that affects a large number of students in the U.S. [12].

A. Setting and participants

The study included five instructors at four geographically disparate U.S. institutions. Table I provides a summary of the characteristics of the participating instructors, courses, and schools. Each course fulfills some portion of its respective institution's requirements for general education science content; accordingly, most of the participating students are nonscience majors. Two courses (those taught by Nick and Shawn, pseudonyms) are "second tier" or upper division; however, these are still introductory level and satisfy general education requirements, and there is no prerequisite of another astronomy course. The other three instructors' (i.e., Olga, Seth, and Jeff; also pseudonyms) courses are considered introductory (100 level).

The courses' content varies somewhat, but there are notable commonalities. Seth's and Jeff's courses are surveys of the whole Universe, including topics such as the nature of light, our solar system, stars, galaxies, cosmology, and certain physical principles (e.g., gravity, Kepler's laws of planetary motion). This structure is consistent with many introductory astronomy courses [16]. Courses taught by Olga and Nick are "stars and galaxies"—in practice these courses cover the same topics listed previously except the solar system. Shawn's course is called simply "stars" and focuses on this topic in greater

TABLE I. Participating instructor, course, and school characteristics.

Instructor name ^a	Approximate course enrollment	N^b	Percent female	Percent prior astronomy coursework	Unique course characteristics	School type ^c
Olga	5 sections of 25 each	71	70.4%	49.3%	Online; "stars and galaxies"	Public, Associate's suburban-serving multi-campus
Nick	1 section of 100	60	26.7%	21.7%	Second-tier (300-level); "stars and galaxies"	Public, Masters-large
Seth	1 section of 800	375	52.0%	7.5%	Concert hall setting; whole universe	Public, research university/very high ^d
Jeff	2 sections of 150 each plus 1 section of 60	100	58.0%	6.0%	Planetarium setting	Public, Masters-large
Shawn	1 section of 600	417	54.4%	40.0%	Concert hall setting; second-tier (200-level); "stars"	Public, research university/very high ^d

^aPseudonyms.

^b N here represents the number of students who consented for their responses to be used in the study.

^cFrom Carnegie Foundation for the Advancement of Teaching, Institutional Classifications (2010 edition, retrieved 11/2/2012 from http://classifications.carnegiefoundation.org/lookup_listings/institution.php).

^dSeth and Shawn teach at the same institution.

depth (though still including relevant topics such as the nature of light). In agreeing to participate in the study, all of the instructors felt the Star Properties Concept Inventory (SPCI) [24,73], described below, adequately assessed a sizable portion of the course content. Several of the instructors had used the SPCI in the past for their own purposes and were willing to use it again. We have made a nonstatistical generalization here [74] that the students in these courses can be grouped together based on our knowledge of the schools and courses as well as the pretest scores (as described further below).

B. Instruments and data collection

We used three instruments to measure participants' interest in selected astronomy topics, self-efficacy toward course tasks, and understanding of star properties. We describe each of these instruments in detail below, presented in the order administered to participating students as a single, combined instrument. Item 1 was a question of informed consent. Although all students in the courses completed the survey, responses from students who declined consent were removed prior to any analysis.

1. Student interest in selected astronomy topics

The first instrument was a short survey to investigate students' interest in three related topics within the course (star properties, stellar evolution, and the Sun) as well as a fourth topic common to introductory astronomy but not specifically measured by the knowledge instrument (the solar system). The purpose of this additional topic was to ascertain whether the participant's interest was focused only on the particular subject of stars or included astronomy more generally. The focus on stars—rather than simply astronomy—was purposeful, in order to both make connections to the knowledge instrument and the fact that one of the participating classes (Shawn's) concentrated on this content area. Participants indicated their interest in each selected topic using a 5-point Likert scale, with 1 representing “not at all interested” and 5 being “very interested.” The survey was adapted from a self-created instrument by Cordova *et al.* [51], which had an original reliability of $\alpha = 0.87$. This instrument is labeled “INT” in the discussion below and is presented in Table II. For the

present study, reliability of INT scores exceeded the acceptable threshold of 0.7 [75], with Cronbach's $\alpha = 0.82$ at preinstruction and 0.87 at postinstruction.

2. Self-efficacy for course tasks

The second instrument aimed to measure participants' self-efficacy about tasks common to general education courses; it was also adapted from Ref. [51] (original $\alpha = 0.94$). The specific tasks selected have a high likelihood of appearing in an introductory astronomy course, such as taking exams or reading about content. The instrument used a 5-point Likert scale to assess students' self-efficacy for course tasks, where 1 was labeled “I cannot do this at all” and 5 was “highly certain that I can do this.” In the discussion below, this instrument is referred to as “SE”; it is shown in Table III. For the present study, reliability of SE scores exceeded the acceptable threshold of 0.7 [75], with Cronbach's $\alpha = 0.94$ at preinstruction, and 0.91 at postinstruction.

3. Student understanding of stars

The SPCI [24,73] measures student understanding about stars. The SPCI contains items relating to star properties (such as mass, temperature, and lifetime), star formation, and the nuclear fusion process that powers stars. The development process showed that the SPCI could measure changes in student understanding about these ideas over an intervention (a unit or course) [24,73]. We used version 3.5 for the present study, containing slight modifications from the previously referenced version as a result of a larger, ongoing study. The SPCI v3.5 contains 22 items presented in multiple-choice format, scored as 1 (correct) or 0 (incorrect), as well as items about gender and prior astronomy coursework. For the present study, reliability of SPCI post scores exceeded the acceptable threshold of 0.7 [75], with Cronbach's $\alpha = 0.78$. For SPCI pre, $\alpha = 0.69$, which is just slightly below the threshold. We consider this lower reliability value acceptable because of the relatively low amount of knowledge typically held by students prior to instruction. Reliability measures are sensitive to homogeneity within the sample [76], and, at the pretest, it is likely that many participants are similar in their lack of knowledge about stars.

TABLE II. Interest instrument, administered as Items 2–5 in the combined survey.

[Instructions] For the following items, choose the number that describes your level of interest in:					
	Not at all interested	Slightly interested	Moderately interested	Interested	Very interested
2. Star properties	1	2	3	4	5
3. Stellar evolution	1	2	3	4	5
4. The Sun	1	2	3	4	5
5. The solar system	1	2	3	4	5

TABLE III. Self-efficacy instrument, administered as items 6–11 on the combined survey.

[Instructions] Please rate how certain you are that you can perform the academic tasks listed below in a class on stars. Rate your degree of confidence by recording a number from 1 (Cannot do at all) to 5 (Highly certain can do).

	I cannot do this at all		Moderately certain I can do this		Highly certain I can do this
6. Understand the readings for a class on stars.	1	2	3	4	5
7. Learn the basic concepts taught in a class on stars.	1	2	3	4	5
8. Understand the material presented by an instructor on stars.	1	2	3	4	5
9. Understand the diagrams and pictures used in a class on stars.	1	2	3	4	5
10. Do well on assignments in a class on stars.	1	2	3	4	5
11. Do well on exams in a class on stars.	1	2	3	4	5

4. Data collection process: Surveys

Students responded to a total of 35 questions (in order informed consent, INT, SE, and SPCI). Administration took place through different methods, depending upon the unique context of the course: (a) Olga, Seth, and Nick gave the survey online, outside of any class meeting time; (b) Jeff administered the survey traditionally, with each student receiving a copy of the survey and a scannable answer sheet; and (c) Shawn displayed the questions in small chunks via electronic slides on the room’s media screen while each student replied on his or her own scannable answer sheet. Administration for methods (b) and (c) took approximately 30 min. Instructors gave or, in the case of method (a), announced the survey on the first day of class, immediately after a brief introduction to the course syllabus. The posttest administration occurred during the classes’ final two weeks in the same format as the pretest. None of the instructors graded the SPCI for correctness, though some gave participation points for completion. Those points were given independently of whether the student consented to be part of the study. Data screening did not raise any concerns about differences in test administration type (paper versus electronic) or effort based upon whether or not the students received participation points; additionally, the instructors did not recall any behavioral issues in the administration that would cause concern over student effort. Hake [66] argued, based on his data set of more than 6000 FCI pretest and posttest pairs from multiple institutions, that grade credit did not seem to influence posttest scores, and that his survey of instructors and personal experience suggested that pretests were similarly taken seriously by participating students.

The first author randomly assigned an identification number to each pretest response, whether electronic or hard copy. Per human subject regulations, respondents’ names were then removed from the data and the first author maintained a secure, confidential list matching names to numbers. We later matched posttest names to the pretest in order to use the same identification number when possible, and again removed names from the data. Next, we used Excel to remove nonconsenting responses and selected for

those participants who completed both the pretest and the posttest. Analyses were based on this matched data set.

5. Instructor interviews

After initial quantitative analyses were performed (described below), we observed differences in self-efficacy for course tasks across instructors that were unexpected and desired to further investigate the possible causes of them. In discussing the quantitative results and our knowledge of the courses themselves, we hypothesized that the instructional design and implementation might relate to the different sources of self-efficacy as described by Bandura [4]. At this point, our focus shifted from the students (their knowledge, self-efficacy for course tasks, and interest as they relate to stars) to the instructors (their instructional design and implementation). We created a retrospective, semistructured interview protocol to elicit information about potential sources of self-efficacy; questions related to the nature of the course, such as the structure of the “lecture” period, assignments, and grading policies, presence and role of teaching assistants, and students’ use of office hours. The first author conducted telephone interviews with each instructor to better understand their instructional design and implementation, as well as similarities and differences across the courses. Interviews were audio recorded and lasted between 20 and 60 min each.

C. Data analysis

To explore our research questions fully, we conducted first quantitative analyses on the student-based interest, self-efficacy for course tasks, and knowledge survey responses and then a qualitative analysis on the instructor interviews. Our quantitative analysis strategy consisted of three major components. First, we screened the data for outliers. We found no multivariate outliers as measured by Mahalanobis distance; however, we removed 15 participants from the analysis because we considered their scores to be representative of univariate outliers (i.e., $z \geq |3|$ in any variable score). We also removed all participants who indicated they had taken astronomy previously. Preliminary analysis showed that those participants who had taken

astronomy before had significantly higher scores on all three pretest variables and therefore were not representative of the larger sample. This resulted in the removal of a little more than 300 students (leaving $n = 692$). At this time, we also checked the data for assumptions of the relevant tests, such as skewness and kurtosis to gauge normality, and examined the descriptive statistics of the sample; discussion of these tests is included in Sec. V.

Second, we performed a sequential multiple regression analysis (i.e., where we conducted incremental partitioning of variance to control for preinstruction conditions, and examined the simple and direct relations between the predictor and the criterion variables) to determine how posttest self-efficacy for course tasks predicted postinstruction understanding (RQ1). At this point in the analysis, we were looking globally at the group of undergraduate astronomy students, and so combined the courses into a single data set. Third, we performed a multivariate analysis of variance (MANOVA) and relevant follow-up tests about how each of our three variables changed over the semester (RQ2a). The follow-up tests also allowed us to look for instructor differences (RQ2b). We have chosen these analyses over looking at normalized gain [66] because of several known problems, such as violations of normality assumptions inherent within ordinary least squares (OLS) analyses and bias toward high pretest scores (for more information, see Refs. [77,78] and references therein). The analyses chosen can further explain more variance than can normalized gain and therefore provide greater statistical power [79].

As described above, the results of the quantitative analyses led us to conduct interviews with the five instructors, which we then analyzed qualitatively. The interviews were transcribed and analyzed by two of the authors (J. M. B. and J. R. C.). We conducted an *a priori* content analysis [80] in which we specifically looked for evidence of the four sources of self-efficacy as described by Bandura [4], as well as any other common themes that might have emerged from the data. We also identified relevant exemplar statements that could support the analysis and interpretations. The coding process followed common techniques as described by Creswell [81] or Seidman [82], such as highlighting in a given color statements made by the instructors that related to a particular source of self-efficacy and then creating a thematic profile of the highlighted statements to verify consistency across the data and the reviewers. The resulting analysis was used by the researchers to inform the interpretation of the self-efficacy differences observed. We discuss all results—both quantitative and qualitative—in the next section.

V. RESULTS

In the next sections, we provide the results of several different analyses of the data collected. We start with descriptive statistics for the sample before moving into the specific approaches to answer the research

questions—sequential multiple regression analysis for understanding how the independent variables (SPCI, SE, and INT pre; SE post; and INT post) predict the dependent variable (SPCI post) and repeated measures MANOVA to understand how the three constructs changed over time. We then discuss the qualitative component of the study—instructor interviews—and how they can enlighten the quantitative results.

A. Descriptive statistics

After the data screening described in Sec. IV.C, we calculated basic descriptive statistics for the three measured variables (INT, SE, and SPCI) at both pre and postinstruction. This allows us to get a general sense of the features of the data; these descriptive statistics are embedded within the inferential statistical analyses as elaborated below. Table IV shows the pre and post means and standard deviations for both individual instructors and an omnibus overall set, and coefficient alpha (i.e., reliability) values on overall, for all three variables. We then calculated bivariate correlations in order to determine the associative relations between the study variables. As shown in Table V, the bivariate correlations reveal significant and positive associations between all the study variables at pre and postinstruction. However, although correlations were all significant, most of the associations represent a small to moderate effect (all Pearson r values ≤ 0.5) [83].

TABLE IV. Descriptive statistics for the study variables.

Variable	n	Preinstruction			Postinstruction		
		M	SD	α	M	SD	α
INT ^a							
Overall	692	3.41	0.881	0.821	3.45	0.934	0.867
Olga	36	3.65	0.860		3.52	1.08	
Nick	47	3.99	0.740		4.00	0.824	
Seth	325	3.37	0.886		3.35	0.931	
Jeff	89	3.54	0.735		3.51	0.835	
Shawn	195	3.23	0.899		3.44	0.935	
SE ^a							
Overall	692	3.86	0.780	0.935	4.03	0.756	0.909
Olga	36	4.02	0.680		3.81	0.836	
Nick	47	4.28	0.660		4.11	0.752	
Seth	325	3.68	0.749		3.99	0.756	
Jeff	89	3.99	0.740		3.84	0.754	
Shawn	195	3.97	0.824		4.19	0.711	
SPCI ^b							
Overall	692	6.81	2.89	0.687	13.0	4.19	0.781
Olga	36	6.72	2.88		11.9	4.38	
Nick	47	9.21	3.24		15.0	3.91	
Seth	325	6.48	2.95		13.7	4.08	
Jeff	89	6.28	2.65		11.3	3.98	
Shawn	195	7.04	2.50		12.3	4.06	

^aINT (interest) and SE (self-efficacy) scores were each measured on a 1 to 5 point Likert scale.

^bSPCI scores were measured out of 22 possible correct.

TABLE V. Correlation Matrix of the Study Variables.

	1	2	3	4	5	6
1. INT ^a -pre	...					
2. INT ^a -post	0.465 ^b	...				
3. SE ^a -pre	0.277 ^b	0.167 ^b	...			
4. SE ^a -post	0.194 ^b	0.340 ^b	0.359 ^b	...		
5. SPCI ^a -pre	0.161 ^b	0.171 ^b	0.254 ^b	0.274 ^b	...	
6. SPCI ^a -post	0.145 ^b	0.231 ^b	0.186 ^b	0.491 ^b	0.366 ^b	...

^aINT = interest, SE = self-efficacy, and SPCI = Star Properties Concept Inventory scores.
^b $p < 0.01$.

B. Sequential multiple regression analysis

We conducted a sequential multiple regression analysis [84] to examine how well self-efficacy for course tasks and interest predicted students’ understanding of star properties after instruction (SPCI post, the criterion variable); results are summarized in Table VI. Sequential multiple regression (sometimes called incremental partitioning of variance) [85] is a method that accounts for each predictor variable’s contribution to overall variance (i.e., R^2). For the purpose of examining our first research question, sequential multiple regression provides a parsimonious initial examination of the predictive relationship and specific contribution of postinstructional self-efficacy for course tasks and interest to the variance in post-instructional knowledge.

In the first step of the sequential multiple regression, SPCI, SE, and INT pretest scores were entered in order to statistically control for the impact of these variables. SE post was entered as the second step and INT post as the third step. The analysis revealed that R^2 for the full regression was significant, $F(5\ 686) = 53.7$, $p < 0.001$, indicating that 27.6% of the variance in SPCI postscores was explained by the predictor variables. SPCI pre, SE pre, and INT pre together accounted for 8.6% of the variance,

TABLE VI. Sequential multiple regression parameters explaining understanding of star properties after instruction (SPCI post).

	R^2	ΔR^2	ΔF	b^c	β^c	Δp^d
Step 1:	0.086	0.086	22.6			
SPCI ^a pre				0.382	0.263	<0.001
SE ^b pre				0.296	0.055	0.149
INT ^b pre				0.369	0.078	0.041
Step 2: SE ^b post	0.276	0.190	181	2.56	0.463	<0.001
Step 3: INT ^b post	0.276	0.002	1.53	0.211	0.047	0.217

^aSPCI scores were measured out of 22 possible correct.
^bSE (self-efficacy) and INT (interest) scores were measured on a 1 to 5 point Likert scale.
^c b and β represent unstandardized and standardized, respectively, weighting coefficients that could be used as a multiplier of an individual’s pretest score to predict their posttest score.
^d Δp is the p value of the partitioned variables.

$p < 0.001$, as demonstrated by the first step. SE post accounted for an additional 19.0% of the variance above and beyond the pretest measures, $p < 0.001$. Interestingly, the variance explained by SE post is more than twice the influence of SPCI pre, SE pre, and INT pre combined, representing a very large effect size and strong practical significance for SE post. INT post did not account for any additional significant variance, $p = 0.22$. Because INT post was an insignificant factor, we did not include this variable in any subsequent analyses. We will revisit this finding in Sec. VII.

C. Repeated measures MANOVA

We also conducted a repeated measures MANOVA, with SPCI and SE scores as the dependent measures, to answer RQ2. This analysis is designed to test group differences where there is more than one dependent variable. The within-subjects variable was time (i.e., pre and postinstruction) and the between-subject variable was instructor. Because of our uneven group sizes across instructors, our analysis did not meet the assumption of equality of covariance matrices (homoscedasticity), with Box’s $M = 139$, $p < 0.001$. Therefore, we used the more stringent Pillai’s criterion in our evaluation of significance testing [84]. The repeated measures MANOVA revealed a significant interaction between time and instructor for the combined scores of SE and SPCI, $F(122\ 061) = 5.99$, $p < 0.001$, $\eta_p^2 = 0.034$. Follow-up repeated measures univariate analyses of variance (ANOVAs) indicated that the interaction between time and instructor was statistically significant for both SPCI, $F(4\ 687) = 9.37$, $p < 0.001$, $\eta^2 = 0.052$, and SE, $F(4\ 687) = 8.26$, $p < 0.001$, $\eta^2 = 0.046$.

1. Simple effects analysis: SPCI

Our follow-up simple effects analyses revealed notable differences between the instructors (see Table IV for variable means and standard deviations by instructor). As shown in Fig. 1, SPCI prescores for Nick’s classes were significantly greater than all the other instructors’ classes (all pairwise comparison p values < 0.01), with no significant difference between the other instructors in SPCI prescores (all p values $\geq .28$). For SPCI postscores, there was no significant difference between Nick’s and Seth’s classes ($p = 0.36$); furthermore, both Nick’s and Seth’s SPCI postscores were significantly greater than Jeff’s and Shawn’s classes (all p values < 0.01). Nick’s postscore was also significantly higher than Olga’s ($p < 0.01$), however, Seth’s was not ($p = 0.11$). Finally, Olga’s, Jeff’s, and Shawn’s SPCI postscores were not significantly different from one another (all $p = 1.0$). Seth’s classes experienced the greatest increase from relatively low preinstruction scores to levels that were statistically equivalent to the highest SPCI scores at postinstruction.

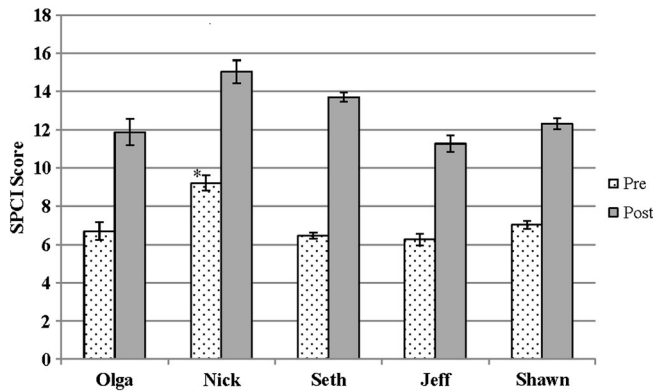


FIG. 1. Mean SPCI scores by instructor. All postscores are significantly greater than prescores ($p < 0.001$). *Nick's mean prescore was significantly greater than all other instructors' scores ($p < 0.01$). Bars show standard errors.

2. Simple effects analysis: SE

As shown in Fig. 2, classes taught by two of the instructors showed significant increases in SE scores over the semester (Seth, $p < 0.001$, and Shawn, $p < 0.001$), whereas the others had no significant change in SE (all p values > 0.12). At preinstruction, Seth's class's SE prescores were significantly lower than the classes taught by Nick, Jeff, and Shawn (all p values < 0.01); SE prescores for Olga's and Seth's classes were not significantly different from one another ($p = .13$). On the other hand, SE post scores for Seth's classes were not significantly different from the other instructors' classes (all p values = 1.0), except Shawn's class. In this case, Shawn's class had significantly greater SE postscores than Seth's class ($p = 0.033$). This showed that Seth's class, who had significantly lower SE scores at preinstruction, experienced the greatest change in SE scores—leading to scores that equaled or significantly exceeded the other instructors'

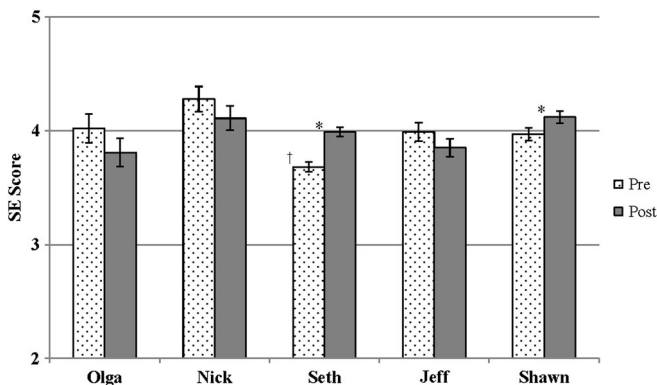


FIG. 2. Mean SE scores by instructor. *Seth's and Shawn's posttest scores were significantly higher than their pretest scores ($p \leq 0.001$). †Seth's pretest score was significant lower than all other instructors' (all $p < 0.01$), except Olga's ($p = 0.13$). Bars show standard errors.

classes at postinstruction except Shawn, who also had an increase.

D. Instructor interviews

As indicated by the results of the repeated measures MANOVA, student self-efficacy for course tasks significantly increased over the course of instruction under two of the five instructors and was correspondingly related to larger gains in knowledge. Therefore, the lead author conducted semistructured phone interviews with all five of the instructors to try to understand what differences in their course design and implementation might help explain the variation in student self-efficacy for course tasks. The first author asked instructors to provide more information regarding their course design, classroom environment, student opportunities to practice performing the tasks necessary for course success, the opportunities students had to witness others like them performing successfully in the course, and the feedback and verbal persuasion students received.

Many similarities existed among the instructors' courses. For instance, most of the instructors (Nick, Seth, Jeff, Shawn) relied heavily on lecture, in-class collaborative lecture tutorials [71], and conceptual questions in a think-pair-share [86] or peer instruction [87,88] model, which are designed to facilitate student discussion and reflection on course content. However, Olga's class was taught entirely online and asynchronously, and so had very different structures in place (e.g., no lectures but more reading of the textbook and accompanying notes). All five instructors discussed attempts to create a positive class environment and most discussed their attempts to increase student interaction and motivation, as well as student responsibility for their own learning. For example, Shawn described his first day of class in this way:

I tell them that I will meet them halfway...What I promise is that I will make this class accessible. I'm not going to make it easy because that's a waste of time... I am going to explain things in terms that you can understand so that you can succeed in this class, even if you have poor math skills. In return, I ask you to participate. Come to every class, do the assignments on time, ask questions. If you participate fully, you will learn something from this class and you'll also get a decent grade. It's my job to make sure I make this class accessible to you. Your job is to participate. So I ask them if they will enter into that agreement with me.

However, there were some notable differences between the two instructors (Seth, Shawn) who saw significant increases in student self-efficacy for course tasks scores and the other three instructors (Olga, Nick, Jeff). The differences identified directly relate to the main factors theorized to impact self-efficacy [4]. All of the instructors

indicated that they tried to encourage and motivate students with their words in addition to providing opportunities to have mastery experiences with the course content. However, the degree of mastery experiences, type of verbal persuasion provided, and the opportunities students had for vicarious experiences of success seemed to differ. This difference may have, in turn, accounted for the variation in self-efficacy for course tasks between the instructors' classes. We should note that our interviews provided little evidence of the influence of physiological factors on potential changes in self-efficacy for course tasks, as this source is based upon the students' interpretation of their own physiological response rather than the instructor's perception of their response. Therefore, in the following two sections, we discuss differences in the other three sources of self-efficacy for course tasks among the instructors' classes.

1. Mastery experiences and verbal persuasion

Instructors frequently discussed feedback to students in the interviews, however, they infrequently differentiated between performance-related feedback provided on a mastery experience (e.g., "this answer is correct or incorrect") and more general feedback provided as verbal persuasion (e.g., "I know you guys can do this"). As a result, in the subsequent discussion we will highlight feedback as mastery or persuasion if the inference was clear to us; in other cases, we will simply use the generic term feedback.

The two instructors (Seth, Shawn) whose students demonstrated greater gains in self-efficacy for course tasks, and subsequently knowledge, appear to provide a combination of more effective scaffolding (i.e., instructor-controlled aspects of the task that are initially beyond students' capabilities and facilitate students' successful task completion [89]) and greater opportunities for practice and feedback than the other instructors. Based upon the descriptions of their courses, Seth and Shawn give more specific direction regarding the actions students need to perform to master the content and do well in the course, as well as provide more purposefully planned opportunities for students to perform those actions. The students also receive elaborated feedback on their performance, even before the first exam or quiz. Feedback may help students recognize and resolve inconsistencies between their actual performance and the target performance [90]. Because of the large class size, such feedback was not necessarily at the individual level but instead more often involved large-scale, class-based discussion of the in-class activities and think-pair-share or peer instruction questions. It is possible that the multiple performance opportunities coupled with ongoing extensive performance-related feedback and additional scaffolding may have helped students of Seth and Shawn become more metacognitively aware of their actual level of knowledge (as compared to the target knowledge)

and the learning strategies necessary to succeed. The repeated opportunities for practice and accompanying performance-related feedback likely helped to increase the students' abilities to gauge the accuracy and completeness of their knowledge and adjust their study skills accordingly, as well as increase the likelihood for future mastery experiences over the course.

Seth illustrated this point through his discussion of the strong emphasis he puts on encouraging students to not just know the right answer when it comes to the think-pair-share questions, lecture tutorials, etc. (practices that are common in four of the five instructors' course lectures), but to be able to reason through and solve the problems both orally and in writing. He discussed the many opportunities the students had to practice or demonstrate the necessary skills in class and to receive performance-related feedback. He also stated that they had "repeated opportunities to be metacognitive, to assess their own understanding, to gain feedback through whether or not they are able to answer think-pair-share questions, or fill-in-the-blank in real time, or make a prediction in real time." He stated that he purposely scaffolds the think-pair-share questions starting with the easier ones to build confidence, and eventually leads up to the harder ones that then lead into a lecture tutorial. Shawn also discussed the importance of practice experiences and feedback in his class. As well as the opportunities in class, he also gave students opportunities to reason through problems in both writing and orally through homework assignments, which were closely aligned with his weekly quizzes (used in lieu of less frequent, high-stakes exams). Shawn also stated that he provided extensive individual feedback on a first draft of a research paper, as well as automated relevant feedback on the homework assignments. Therefore, classroom and homework activities in both Seth's and Shawn's classes supported the development of self-efficacy for course tasks through mastery experiences.

The three instructors who did not see increases in student self-efficacy for course tasks, however, gave exams or quizzes only, or exams or quizzes in conjunction with homework assignments that received little or no performance-related feedback and were not scored for correctness. For instance, Nick said that he emphasizes to his students "over and over again the extent to which these group activities and their time spent outside of class, which I don't know how much that is, but that the time they spend going over these materials and studying them is the only way they can learn." However, his students are given no formal homework assignments to practice the necessary skills and no feedback beyond exam scores unless they meet with him during office hours. Olga, who teaches online and has limited resources in terms of time and assistance, indicated that students have limited opportunities to practice the necessary tasks, and do not receive feedback. Therefore, in Olga's, Nick's, and Jeff's classes,

students were provided fewer opportunities to engage in mastery experiences than in Seth's and Shawn's classes and received less feedback accordingly.

The two instructors (Seth, Shawn) whose students achieved greater gains in self-efficacy for course tasks and higher knowledge made no mention of student failure experiences. Interestingly, however, the other three instructors all mentioned student failure as a common theme at the first exam or quiz. With fewer opportunities for mastery experiences and feedback before the first exam, and therefore little metacognitive awareness of the discrepancies between their perceived and actual knowledge upon taking the first exam, exam failure was more salient. These instructors indicated that their students are typically surprised when they perform poorly. For instance, Olga mentioned that students often get discouraged after the first quiz because their scores are lower than they are used to getting in other classes; Jeff said that students find the first exam "very shocking" and that "the normal mode for the average student is they tend to do poorly on the first exam".

Like all the instructors, Jeff tries to provide verbal encouragement to the students, letting them know that if they are willing to change their behavior and put in the effort, they can do better. He stated that after the first exam, the students are "angry because they feel they studied really hard and yet they didn't succeed," therefore he has what he calls a "defusing the bomb day." During this meeting, he shows them where the commonly missed items were covered in the previous class materials to show them "how the exam linked back to the activities we did" and that they "could have gotten this question correct." Whereas the intention to make students responsible for their own learning (through mastery experiences) and to encourage them by letting them know that they are capable of performing successfully with adequate effort (through verbal persuasion) may be good, it is possible that pointing out the students' failure and implied lack of effort, when the students feel they did put forth effort, may not have had a positive effect on student self-efficacy for course tasks and may have instead been quite discouraging.

2. Vicarious experiences

Another difference may relate to the employment of undergraduate teaching assistants as peer models of success (used in courses taught by Seth, Shawn, and Jeff), which increased the degree of vicarious experiences available to students in these classes. In all three cases, the undergraduate teaching assistants were students who had previously taken the introductory course, performed well, and demonstrated an interest in helping others. Both of the instructors whose students exhibited increases in self-efficacy for course tasks used undergraduate teaching assistants in their courses who may have been perceived as peer models of success by the students. Although one of

the three instructors (Jeff) without increases in student self-efficacy for course tasks also used undergraduate assistants, student teaching assistants may have played a much larger role in the courses (Seth's and Shawn's) where student self-efficacy for course tasks increased. The teaching assistants in Jeff's course (i.e., one without a significant increase in self-efficacy for course tasks), for example, assisted with in-class lecture tutorials and did minor clerical work. However, in addition to assisting in class, the student teaching assistants in the other two courses (Seth's and Shawn's) held regular office hours for students, providing another avenue for assistance and interaction with peer models of success.

Were undergraduate teaching assistants regarded by the students as peers? In other words, did students view undergraduate teaching assistants vicariously, or did students perceive them as experts (i.e., in much the same way as a graduate teaching assistant or even the professor is and so cannot be a vicarious source of self-efficacy)? We do not have data in this study (e.g., from interviews with students on their interpretation of the undergraduate assistants' role) to know for certain. However, the ways in which both Seth and Shawn refer to the assistants heavily emphasizes that they were very recently in the same position as the students—they completed this same course only the semester or year before, struggled in the same ways as the current students will, and so forth. As such, although their support may also contribute to students' mastery opportunities, the structure is designed for them to be seen as knowledgeable peers rather than experts and so their contributions to vicarious experiences would be expected to dominate.

3. Summary of qualitative analysis

The main differences between the courses where students demonstrated increases in self-efficacy for course tasks and those that did not seem to be in more opportunities for mastery experiences (e.g., through scaffolded instruction); more extensive, purposeful performance-related feedback on those experiences; differences in the type and timing of verbal persuasion offered; and more extensive use of vicarious experiences (i.e., successful peer models). One possible alternate explanation could be class size: the two instructors whose students demonstrated increases in self-efficacy for course tasks happened to teach extremely large courses (600+), whereas the other three instructors' courses ranged from under 35 up to about 160 students per section. However, we feel this is unlikely to be related to gains in self-efficacy over the course because common sense would lead us to believe that larger classes would result in feelings of anonymity that could, in turn, reduce self-efficacy. Another possible explanation could be the nature of the university and/or students. For instance, two of the three instructors who did not experience significant increases in their students' self-efficacy for course tasks mentioned the unique nature of their students,

although there is little evidence to support their claims at least with respect to the variables measured in this study (i.e., their students' preinstruction scores were not significantly different from those at other institutions). Jeff indicated that the university he teaches at is more of a comprehensive college or open enrollment university with a "half community college, half university crowd" of students. Olga also stated that her students were a mix of associate-level students and those transferring to four-year universities, with many nontraditional students. However, we do not believe the nature of the students can account for the differences, as Jeff's students exhibited not only significant gains in knowledge over the course but also relatively high self-efficacy for course tasks at preinstruction. We have no reason to believe that student effort made a difference in pretest scores, which are (with the exception of Nick's students in knowledge) statistically the same, as no class had a high level of nonparticipation vis-à-vis incomplete or nonsense responses.

VI. DISCUSSION

This study investigated the impact of students' self-efficacy for course tasks and interest on their knowledge about stars in undergraduate general education astronomy courses. Both of these factors have been included in conceptual change models such as the CRKM [1] as part of personal relevance, which may result "from interest, emotional involvement, [or] self-efficacy" [3], (p. 110). Using sequential multiple regression, we found that the combined preinstructional variables of prior knowledge about stars (SPCI scores) and self-efficacy for course tasks explained a significant amount of the variance in postinstruction knowledge about stars—however, self-efficacy for course tasks postscores explained more than twice as much of the variance in postinstruction knowledge than the preinstruction variables combined. We followed up the sequential regression with a repeated-measures multivariate analysis of variance (MANOVA) to examine potential differences between instructors. Whereas all instructors' classes demonstrated an increase in star knowledge from pre to postinstruction, only two of the instructors' classes showed increases in self-efficacy for course tasks. Furthermore, the greatest gains in knowledge were related to this increased self-efficacy for course tasks. This finding supports earlier research that indicates self-efficacy is related to motivation, engagement, achievement, and the use of metacognitive and self-regulatory strategies, as well as the effort and persistence a learner exhibits when faced with a difficult task [4,91–94]. The role of self-efficacy in conceptual change within the CRKM remains an area in need of study. As hypothesized by the CRKM, self-efficacy may have facilitated deeper engagement with the topic, thereby increasing the likelihood for knowledge change [1,3]. Furthermore, increases in self-efficacy may

promote persistence in engaging with the topic further [37], which in turn could deepen understanding.

Based on our qualitative analyses of instructor interviews, it seems that specific instructional practices may help explain the differences in self-efficacy for course tasks postinstruction and subsequent differences in knowledge gains among instructors' students. In addition to the extensive use of peer models of success, which may increase vicarious experiences, these instructional practices include incorporating more purposefully planned opportunities (both formal and informal) for students to practice the skills necessary to succeed in the course (mastery experiences) and receive more detailed feedback regarding those performances throughout the course (particularly before any major assessments such as exams take place). In particular, it seems that continual opportunities to perform the necessary actions for success while receiving relevant, ongoing, elaborated feedback may have helped students become more metacognitively aware of discrepancies between their perceived level of knowledge and actual level of knowledge while simultaneously continuing to practice the skills and improve their performance. It is likely that with ongoing practice and feedback, students were better able to gauge their actual knowledge or performance level and adjust learning strategies accordingly, as well as increase their prospects for mastery experiences throughout the course, which may have had a positive impact on their self-efficacy for course tasks, and subsequently, their content knowledge.

Levels of interest did not significantly explain any of the variance in star knowledge scores at postinstruction. This apparent disconnect seems contrary to perspectives that theorize the importance of interest in conceptual change (see, for example, Ref. [1]). However, our lack of a significant finding may be due to a limitation in this study, where we did not resolve levels of interest into finer-grained categories. Linnenbrink-Garcia and colleagues found no significant relationship between interest and conceptual change when participants were combined into an aggregate whole [95]. It was only when these researchers divided participants into specific levels of interest (high, medium, and low) and specific levels of background knowledge (also high, moderate, and low) were they able to find a meaningful connection between interest and conceptual change, and only at certain combination levels (i.e., high interest and moderate background knowledge).

VII. LIMITATIONS

This study was conducted in five learning situations across multiple instructors and institutions. Although we have argued above that the participants and setting are sufficiently similar that we can combine the data from the courses, there may have been differences—whether pre or postinstructional—not documented here that could influence the outcomes of this study. For example, some of the

participating instructors (Jeff and Olga) felt that their student populations might be different from what they perceived as typical, but no such differences were noted in the preinstructional variables (except for knowledge in Nick's class). None of the instructors noticed any behavioral anomalies, nor were there noticeable differences between classes in terms of responses that needed to be removed for nonparticipation (e.g., "Christmas treeing" the survey), that might suggest that a given class put forth less effort than the others. Another potential confound relates to the method of administration of the data collection instruments. We do not believe this to be a factor given the close preinstructional scores, but we cannot rule out the possibility of complications as a result of these differences.

Some might argue that the SPCI, which includes some items that are low-level knowledge questions, can measure only knowledge construction rather than conceptual change. Whereas we agree that individual items cannot do this, the SPCI—like many other concept inventories—is specifically designed to measure changes about a single or narrow group of related concepts. For example, Lindell, Peak, and Foster identify different concept inventories and evaluate their design methodology [63]. In doing so, they note that 11 of the 12 instruments whose detailed design information was available for analysis contain three or fewer concepts (as determined by Lindell, Peak, and Foster). Other data collection methods, such as interviews or open-ended questions, would allow more detailed insight into students' understanding of stars, and such techniques were used in the SPCI development process [24,73]. Although we believe the SPCI to be an appropriate measure for the study, we acknowledge that we cannot differentiate at this time whether students experienced deep conceptual change, or a more ephemeral change in knowledge.

Our original hypotheses included that interest could change over an introductory astronomy course, and that interest could be a predictor of changes in knowledge as suggested by the CRKM [1]. We found no significant changes in interest nor any significant variance in post-instructional SPCI scores explained by interest. One possible reason for this may be the relatively high (~ 4 on a 5-point scale) scores at pretest, particularly because all of the courses satisfy general education requirements for science and are therefore chosen by the students over other science topics. A different data collection strategy—whether through a modified or alternative instrument or through interviews with students at various points in the semester—might provide different insights than what we observed here, and should be considered for future study.

Identifying the possible causes of changes to self-efficacy is limited by the data collection, i.e., postinstructional interviews with the instructors in which they self-reported on their course design and instructional strategies. It was not possible to make observations of the course or collect data from students about these aspects.

However, the first author had previous experience observing in some of the instructors' classrooms, and felt that the instructors' responses to interview questions were reasonable reflections of similar past courses. A more in-depth study of courses, including perhaps an ethnographic approach to studying the instructional environments, could provide verification or refutation of our interpretations from interview data. Additionally, the inferences made about sources of self-efficacy assume that they apply equally to men and women, as looking at gender differences was beyond the scope of this study. Future research should include opportunities to differentiate between gender to determine whether the differences observed in, for example, physics courses [37] and STEM career persistence [96,97] are seen in introductory astronomy as well.

We argued in Sec. IV.C. that the analysis method (specifically, sequential multiple regression analysis) chosen here has advantages over normalized gain [66] and better explains the variance of scores. If you were to look at normalized gain, you would find that Nick's class has the second-highest normalized gain score, on par with Seth's and noticeably higher than those of Shawn, Jeff, and Olga. This might suggest that Nick was able to improve knowledge scores without a corresponding increase in self-efficacy. Recall, however, that normalized gain is biased toward higher pretest scores, which Nick had (i.e., his class's pretest score was significantly higher than all the others). It therefore is not surprising that Nick's class had a higher normalized gain but not a corresponding significant improvement in the sequential multiple regression analysis.

VIII. IMPLICATIONS FOR INSTRUCTION AND FUTURE RESEARCH

Traditionally, self-efficacy has not been a major consideration in the development of undergraduate science courses, especially those taught in large-lecture format. We have already seen that active engagement curricular materials and instructional strategies improve learning gains over traditional lecture courses. Some of these materials provide opportunities for students to engage in mastery experiences; however, it may be necessary to explicitly identify these opportunities to students. Additionally, elaborate, performance-related feedback on the mastery experiences may be critical for such materials to reach their full potential; if students do not know to what extent they are succeeding or need more work, their self-efficacy may not be improved as much as it might otherwise be. Closely related to this is the broader source of verbal persuasion. Our results suggest that incorporating positive verbal persuasion early and often—especially prior to major assessments such as exams—may improve upon self-efficacy. Linking mastery experiences with performance-related feedback and more general verbal persuasion may provide "more bang for the buck." Finally, vicarious experiences, such as the successful peer models

used in some of this study's courses, may also be beneficial in promoting higher self-efficacy.

The inclusion of self-efficacy, as part of motivational characteristics of the learner in a warm conceptual change model such as the CRKM, may provide some insight into the mechanisms by which students' knowledge changes are facilitated. This kind of research, which goes beyond the simple identification of "what works" and instead tries to better understand the cognitive factors that influence deeper learning is lacking in current undergraduate astronomy education research [10]; for an example in physics, see Ref. [62]. There may also be implications here for faculty professional development, whereby faculty may learn how to better attend to characteristics such as self-efficacy in order to optimize their students' changes in knowledge. Finding ways to bolster self-efficacy in introductory science courses, especially those for nonscience majors and which may include many future teachers [13], is a worthy goal.

Additional research is needed to better understand the role of self-efficacy in knowledge construction and reconstruction. For example, future studies could investigate this issue with a variety of different topics, which may also contribute to the accumulating evidence for domain specificity of self-efficacy [5]. Although it did not make a difference in predicting postinstruction knowledge in this study, understanding more about students' interest in astronomy may be worthwhile given the large population of students in such courses [12] and the increasing attention on professional development for astronomy faculty (e.g., Refs. [98,99]). Improving our own models of the knowledge change processes will ultimately allow us to better facilitate deep understanding for our students.

ACKNOWLEDGMENTS

We would like to thank the five instructors, Olga, Nick, Jeff, Seth, and Shawn—as well as their students—for their time in participating in this study.

-
- [1] J. A. Dole and G. M. Sinatra, Reconceptualizing change in the cognitive construction of knowledge, *Educ. Psychol.* **33**, 109 (1998).
 - [2] P. R. Pintrich, R. W. Marx, and R. A. Boyle, Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change, *Rev. Educ. Res.* **63**, 167 (1993).
 - [3] G. M. Sinatra, The warming trend in conceptual change research: The legacy of Paul R. Pintrich, *Educ. Psychol.* **40**, 107 (2005).
 - [4] A. Bandura, *Self-Efficacy: The Exercise of Control* (W. H. Freeman and Co., New York, NY, 1997).
 - [5] F. Pajares, Self-efficacy beliefs in academic settings, *Rev. Educ. Res.* **66**, 543 (1996).
 - [6] D. H. Schunk, Metacognition, self-regulation, and self-regulated learning: Research recommendations, *Educ. Psychol. Rev.* **20**, 463 (2008).
 - [7] S. Hidi and K. A. Renninger, The four-phase model of interest development, *Educ. Psychol.* **41**, 111 (2006).
 - [8] A. Bandura, Social cognitive theory: An agentic perspective, *Ann. Rev. Psych.* **52**, 1 (2001).
 - [9] J. M. Bailey and T. F. Slater, A review of astronomy education research, *Astron. Educ. Rev.* **2**, 20 (2003).
 - [10] J. M. Bailey, Astronomy education research: Developmental history of the field and summary of the literature (a commissioned paper for the National Research Council Board on Science Education's Committee on the Status, Contributions, and Future Directions of Discipline Based Education Research), 2011 (unpublished), http://sites.nationalacademies.org/DBASSE/BOSE/DBASSE_080124.
 - [11] A. Lelliott and M. Rollnick, Big ideas: A review of astronomy education research 1974–2008, *Int. J. Sci. Educ.* **32**, 1771 (2010).
 - [12] A. Fraknoi, Enrollments in astronomy 101 courses: An update, *Astron. Educ. Rev.* **1**, 121 (2001).
 - [13] F. Lawrenz, D. Huffman, and K. Appeldoorn, Enhancing the instructional environment: Optimal learning in introductory science, *J. Coll. Sci. Teach.* **34**, 40 (2005).
 - [14] G. L. Deming and B. Hufnagel, Who's taking ASTRO 101?, *Phys. Teach.* **39**, 368 (2001).
 - [15] B. Partridge and G. Greenstein, Goals for "Astro 101": Report on workshops for department leaders, *Astron. Educ. Rev.* **2**, 46 (2003).
 - [16] T. F. Slater, J. P. Adams, G. Brissenden, and D. Duncan, What topics are taught in introductory astronomy courses?, *Phys. Teach.* **39**, 52 (2001).
 - [17] S. Vosniadou, Conceptual change research: An introduction, in *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, Taylor, & Francis, New York, NY, 2008).
 - [18] G. J. Posner, K. A. Strike, P. W. Hewson, and W. A. Gertzog, Accommodation of a scientific conception: Toward a theory of conceptual change, *Sci. Educ.* **66**, 211 (1982).
 - [19] M. T. H. Chi, Conceptual change within and across ontological categories: Examples from learning and discovery in science, in *Cognitive Models of Science: Minnesota Studies in the Philosophy of Science*, edited by R. Giere (University of Minnesota Press, Minneapolis, MN, 1992).

- [20] C. A. Chinn and W. F. Brewer, The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction, *Rev. Educ. Res.* **63**, 1 (1993).
- [21] K. A. Strike and G. J. Posner, A revisionist theory of conceptual change, in *Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice*, edited by R. A. Duschl and R. J. Hamilton (State University of New York, New York, NY, 1992).
- [22] W. F. Brewer, Naïve theories of observational astronomy: Review, analysis, and theoretical implications, in *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, Taylor, & Francis, New York, NY, 2008).
- [23] S. Vosniadou and W. F. Brewer, Mental models of the day/night cycle, *Cogn. Sci.* **18**, 123 (1994).
- [24] J. M. Bailey, B. Johnson, E. E. Prather, and T. F. Slater, Development and validation of the Star Properties Concept Inventory, *Int. J. Sci. Educ.* **34**, 2257 (2012).
- [25] E. M. Bardar, E. E. Prather, K. Brecher, and T. F. Slater, Development and validation of the Light and Spectroscopy Concept Inventory, *Astron. Educ. Rev.* **5**, 103 (2007).
- [26] S. D. Hornstein, D. Duncan, and the Collaboration of Astronomy Teaching Scholars, Development of a solar system concept inventory, in *Proceedings the 213th Meeting of the American Astronomical Society*, Long Beach, CA, 2009.
- [27] R. S. Lindell and J. P. Olsen, Developing the lunar phases concept inventory, *Proceedings of the Physics Education Research Conference, Boise, ID, 2002* (AIP, Melville, NY, 2002).
- [28] E. M. Bardar, First results from the Light and Spectroscopy Inventory, *Astron. Educ. Rev.* **6**, 75 (2007).
- [29] E. Brogt, D. Sabers, E. E. Prather, G. L. Deming, B. Hufnagel, and T. F. Slater, Analysis of the Astronomy Diagnostic Test, *Astron. Educ. Rev.* **6**, 25 (2007).
- [30] M. C. LoPresto and S. R. Murrell, Using the Star Properties Concept Inventory to compare instruction with lecture tutorials to traditional lectures, *Astron. Educ. Rev.* **8**, 010105 (2009).
- [31] E. E. Prather, A. L. Rudolph, G. Brissenden, and W. M. Schlingman, A national study assessing the teaching and learning of astronomy. Part I. The effect of interactive instruction, *Am. J. Phys.* **77**, 320 (2009).
- [32] W. J. Straits and R. R. Wilke, Activities-based astronomy: An evaluation of an instructor's first attempt and its impact on student characteristics, *Astron. Educ. Rev.* **2**, 46 (2003).
- [33] S. H. Broughton, G. M. Sinatra, and E. M. Nussbaum, "Pluto has been a planet my whole life!" Emotions, attitudes, and conceptual change in elementary students learning about Pluto's reclassification, *Res. Sci. Educ.* **43**, 529 (2013).
- [34] J. R. Cordova, S. H. Broughton, and G. M. Sinatra, Self-efficacy, confidence judgments, and accuracy in conceptual change learning, *Proceedings of the Annual Meeting of the American Educational Research Association*, New Orleans, LA, 2011; <http://www.aera.net/Publications/Online-Paper-Repository/AERA-Online-Paper-Repository/Owner/429929>.
- [35] J. R. Cordova, G. M. Sinatra, S. H. Broughton, G. Taasobshirazi, and D. Lombardi, Confidence in prior knowledge, self-efficacy, interest and prior knowledge: Influences on conceptual change, *Contemp. Educ. Psychol.* **39**, 164 (2014).
- [36] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, *Psych. Rev.* **84**, 191 (1977).
- [37] V. Sawtelle, E. Brewere, and L. H. Kramer, Exploring the relationship between self-efficacy and retention in introductory physics, *J. Res. Sci. Teach.* **49**, 1096 (2012).
- [38] D. H. Schunk, P. R. Pintrich, and J. L. Meece, *Motivation in Education: Theory, Research, and Applications*, 3rd ed. (Pearson Education, Upper Saddle River, NJ, 2008).
- [39] S. L. Britner and F. Pajares, Sources of science self-efficacy beliefs of middle school students, *J. Res. Sci. Teach.* **43**, 485 (2006).
- [40] S. L. Britner, Motivation in high school science students: A comparison of gender differences in life, physical, and Earth science classes, *J. Res. Sci. Teach.* **45**, 955 (2008).
- [41] H. S. Fencil and K. R. Scheel, Pedagogical approaches, contextual variables, and the development of student self-efficacy in undergraduate physics courses, in *2003 physics education research conference: AIP conference proceedings*, edited by J. Marx, S. V. Franklin, and K. Cummings (American Institute of Physics, Melville, NY, 2004), Vol. 720, pp. 173–176.
- [42] V. Sawtelle, E. Brewere, R. M. Goertzen, and L. H. Kramer, Identifying events that impact self-efficacy in physics learning, *Phys. Rev. ST Phys. Educ. Res.* **8**, 020111 (2012).
- [43] R. E. Bleicher, Revisiting the STEBI-B: Measuring self-efficacy in preservice elementary teachers, *School Sci. Math.* **104**, 383 (2004).
- [44] A. Lakshmanan, B. P. Heath, A. Perlmutter, and M. Elder, The impact of science content and professional learning communities on science teaching efficacy and standards-based instruction, *J. Res. Sci. Teach.* **48**, 534 (2011).
- [45] D. H. Palmer, Sources of self-efficacy in a science methods course for primary teacher education students, *Res. Sci. Educ.* **36**, 337 (2006).
- [46] I. M. Riggs and L. G. Enochs, Toward the development of an elementary teacher's science teaching efficacy belief instrument, *Sci. Educ.* **74**, 625 (1990).
- [47] B. L. Carter, Master's thesis, University of Maryland, College Park, MD, 2005.
- [48] S. J. Slater, T. F. Slater, and A. J. Shaner, Impact of backwards faded scaffolding in an astronomy course for pre-service elementary teachers based on inquiry, *J. Geosci. Educ.* **56**, 408 (2008).
- [49] L. G. Enochs and I. M. Riggs, Further development of an elementary science teaching efficacy belief instrument: A preservice elementary scale, *Proceedings of the National Association for Research in Science Teaching Annual Meeting* (Atlanta, GA, 1990).
- [50] B. Hufnagel, Development of the astronomy diagnostic test, *Astron. Educ. Rev.* **1**, 47 (2002).
- [51] J. R. Cordova, G. M. Sinatra, S. H. Broughton, and G. Taasobshirazi, Self-efficacy, confidence in prior knowledge, and conceptual change, *Proceedings of the Annual*

- Meeting of the American Educational Research Association* (Denver, CO, 2010); <http://www.aera.net/Publications/Online-Paper-Repository/AERA-Online-Paper-Repository/Owner/276919>.
- [52] K. A. Renninger, Interest and identity development in instruction: An inductive model, *Educ. Psychol.* **44**, 105 (2009).
- [53] B. C. Heddy and G. M. Sinatra, Transforming misconceptions: Using transformative experience to promote positive affect and conceptual change in students learning about biological evolution, *Sci. Educ.* **97**, 723 (2013).
- [54] G. M. Sinatra and L. Mason, Beyond knowledge: Learner characteristics influencing conceptual change, in *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, Taylor & Francis, New York, NY, 2008).
- [55] S. M. Reis, G. C. Atamian, and J. S. Renzulli, The effectiveness of a self-instructional curricular unit in the development of advanced level concepts in astronomy, *Gifted Child Quarterly* **29**, 151 (1985).
- [56] R. Jarman and L. McAleese, Physics for the star-gazer: Pupils' attitudes to astronomy in the northern Ireland science curriculum, *Phys. Educ.* **31**, 223 (1996).
- [57] M. Krstovic, L. Brown, M. Chacko, and B. Trinh, Grade 9 astronomy study: Interests of boys and girls studying astronomy at Fletcher's Meadow Secondary School, *Astron. Educ. Rev.* **7**, 18 (2008).
- [58] L. S. Shore, Ed.D. thesis, Boston University, Boston, MA, 1991.
- [59] C. A. Rutkowski, M.Ed. thesis, University of Manitoba, Winnipeg, Manitoba, Canada, 1991.
- [60] T. F. Slater, Ph.D. thesis, University of South Carolina, Columbia, SC, 1993.
- [61] J. M. Bailey, Concept inventories for ASTRO 101, *Phys. Teach.* **47**, 439 (2009).
- [62] G. Taasoobshirazi and G. M. Sinatra, A structural equation model of conceptual change in physics, *J. Res. Sci. Teach.* **48**, 901 (2011).
- [63] R. S. Lindell, E. Peak, and T. M. Foster, Are they all created equal? A comparison of different concept inventory development methodologies, in *2007 physics education research conference: AIP conference proceedings*, edited by L. McCullough, L. Hsu, and P. Herron, (American Institute of Physics, Melville, NY, 2007), Vol. 883, pp. 14–17.
- [64] J. L. Docktor and J. P. Mestre, Synthesis of discipline-based education research in physics, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020119 (2014).
- [65] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- [66] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- [67] R. Duit, D. F. Treagust, and A. Widodo, Teaching science for conceptual change: Theory and practice, in *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, Taylor & Francis, New York, NY, 2008).
- [68] J. T. Leach and P. H. Scott, Teaching for conceptual understanding: An approach drawing on individual and sociocultural perspectives, in *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, Taylor & Francis, New York, NY, 2008).
- [69] D. Jonassen, Model building for conceptual change, in *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, Taylor & Francis, New York, NY, 2008).
- [70] M. C. Linn, Teaching for conceptual change: Distinguish or extinguish ideas, in *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, Taylor & Francis, New York, NY, 2008).
- [71] E. E. Prather, T. F. Slater, J. P. Adams, G. Brissenden, and the Conceptual Astronomy and Physics Education Research (CAPER) Team, *Lecture-Tutorials for Introductory Astronomy*, 2nd ed. (Pearson Addison-Wesley, San Francisco, CA, 2007).
- [72] E. Brogt, A theoretical background on a successful implementation of lecture-tutorials, *Astron. Educ. Rev.* **6**, 50 (2007).
- [73] J. M. Bailey, Development of a concept inventory to assess students' understanding and reasoning difficulties about the properties and formation of stars, *Astron. Educ. Rev.* **6**, 133 (2007).
- [74] G. Keppel and T. D. Wickens, *Design and Analysis: A Researcher's Handbook*, 4th ed. (Prentice Hall, Upper Saddle River, NJ, 2004).
- [75] D. George and P. Mallery, *SPSS for Windows Step By Step: A Simple Guide and Reference, 16.0 Update*, 9th ed. (Pearson Education, Boston, MA, 2009).
- [76] B. Thompson, Understanding reliability and coefficient alpha, really, in *Score Reliability*, edited by B. Thompson (SAGE, Thousand Oaks, CA, 2003).
- [77] L. Bao, Theoretical comparisons of average normalized gain calculations, *Am. J. Phys.* **74**, 917 (2006).
- [78] C. S. Wallace and J. M. Bailey, Do concept inventories actually measure anything?, *Astron. Educ. Rev.* **9**, 010116 (2010).
- [79] E. M. Nussbaum, *Categorical and nonparametric data analysis: Choosing the best statistical technique* (Routledge, New York, NY, 2015).
- [80] S. Stemler, An overview of content analysis, *Pract. Assess. Res. Eval.* **7** (2001), <http://PAREonline.net/getvn.asp?v=7&n=17>.
- [81] J. W. Creswell, *Qualitative Inquiry and Research Design: Choosing among Five Traditions*, 2nd ed. (Sage, Thousand Oaks, CA, 2007).
- [82] I. Seidman, *Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences*, 2nd ed. (Teachers College Press, New York, NY, 1998).
- [83] J. Cohen, A power primer, *Psychol. Bull.* **112**, 155 (1992).
- [84] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 5th ed. (Pearson Education, Boston, MA, 2007).
- [85] J. Cohen and P. Cohen, *Applied Multiple Regression/Correlation Analysis for the Behavioral Science*, 2nd ed. (Erlbaum, Hillsdale, NJ, 1983).
- [86] F. Lyman, Think pair share: The responsive classroom instruction, in *Mainstreaming Digest*, edited by A. S.

- Anderson (University of Maryland College of Education, College Park, MD, 1981).
- [87] P. J. Green, *Peer Instruction for Astronomy* (Prentice Hall, Upper Saddle River, NJ, 2003).
- [88] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).
- [89] D. Wood, J. S. Bruner, and G. Ross, The role of tutoring in problem solving, *J. Child Psychol. Psychiatry* **17**, 89 (1976).
- [90] V. J. Shute, Focus on formative feedback, *Rev. Educ. Res.* **78**, 153 (2008).
- [91] A. Bandura, Adolescent development from an agentic perspective, in *Self-Efficacy Beliefs of Adolescents*, edited by F. Pajares and T. Urdan (Information Age Publishing, Greenwich, CT, 2006).
- [92] E. A. Linnenbrink and P. R. Pintrich, Motivation as an enabler for academic success, *Sch. Psychol. Rev.* **31**, 313 (2002).
- [93] F. Pajares, Gender and perceived self-efficacy in self-regulated learning, *Theory Into Practice* **41**, 116 (2002).
- [94] B. J. Zimmerman and T. J. Cleary, Adolescents' development of personal agency: The role of self-efficacy beliefs and self-regulatory skill, in *Self-efficacy beliefs of adolescents*, edited by F. Pajares and T. Urdan (Information Age Publishing, Greenwich, CT, 2006).
- [95] L. Linnenbrink-Garcia, K. J. Pugh, K. L. K. Koskey, and V. C. Stewart, Developing conceptual understanding of natural selection: The role of interest, efficacy, and basic prior knowledge, *J. Exp. Educ.* **80**, 45 (2012).
- [96] A. L. Zeldin and F. Pajares, Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers, *Am. Educ. Res. J.* **37**, 215 (2000).
- [97] A. L. Zeldin, S. L. Britner, and F. Pajares, A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers, *J. Res. Sci. Teach.* **45**, 1036 (2008).
- [98] G. Brissenden, E. E. Prather, T. F. Slater, W. M. Greene, and M. L. Thaller, The NASA Center for Astronomy Education (CAE): 2007 college astronomy teaching excellence workshops, *Proceedings of the 209th Meeting of the American Astronomical Society*, Seattle, WA, 2007.
- [99] E. E. Prather and G. Brissenden, Development and application of a situated apprenticeship approach to professional development of astronomy instructors, *Astron. Educ. Rev.* **7**, 1 (2008).