


COMMENTARY

Open Access



# A method for analyzing instructors' purposeful modifications to research-based instructional strategies

Erin Scanlon, Brian Zamarripa Roman, Elijah Ibadlit and Jacquelyn J. Chini\* 

## Abstract

**Background:** Numerous studies in the literature describe the effectiveness of research-based instructional strategies (RBIS) in the postsecondary STEM (science, technology, engineering, and mathematics) context. Many of these studies are predicated on the assumption that instructors implement the RBIS exactly as is intended by the developers. However, by necessity, instructors modify the RBIS to suit their needs and to best support their students. The purpose of this commentary is to describe a framework (Modification Identification Framework) and method (Revealed Causal Mapping) for classifying modifications instructors make to an RBIS as they implement it in their course and identify the reasons why instructors make these modifications. As the MIF was developed in the healthcare field, we altered and extended it to be suitable for educational settings. We then demonstrate the usefulness of the framework and method through an extended sample study of instructors' modifications to the Student-Centered Active Learning Environment with Upside-Down Pedagogies (SCALE-UP) model in introductory physics.

**Conclusions:** In general, the findings from investigations with the Modification Identification Framework and Revealed Causal Mapping can be used to identify what experiences lead instructors to modify certain aspects of RBIS. These findings can aid curriculum developers in creating supports for instructors so they can make changes in line with the underlying structure and theory of the RBIS.

**Keywords:** Research-based instructional strategies, Science education, Revealed Causal Mapping

## Introduction

Despite calls to improve postsecondary education and increase the number of degrees awarded in STEM (American Society for Engineering Education, 2012; National Academies of Sciences, Engineering, and Medicine, 2018; National Research Council, 2012; President's Council of Advisors on Science and Technology, 2012) as well as evidence that active learning courses better support student success and conceptual understanding compared to traditional lectures across STEM disciplines (Freeman et al., 2014), instructors still frequently rely on didactic instructional styles (Stains et al., 2018). In fact, studies have demonstrated that one third of physics instructors who try a research-based instructional strategy (RBIS) in their courses discontinue use of all RBIS (Henderson, Dancy, & Niewiadomska-Bugaj,

2012); thus, some instructors attempt to reform their instruction, but ultimately return to more traditional instruction. Similarly, instructors who do continue implementing RBIS often modify the strategy. Foote, Neumeier, Henderson, Dancy, and Beichner (2014) state: "Faculty rarely implement an innovation 'as is', usually adapting ideas to their unique environment, goals, personality, and more. Developers should acknowledge this and focus on helping faculty navigate the difference between productive and unproductive changes, by using change agents or coming up with written recommendations. They should provide advice on how to overcome structural barriers (i.e., budget limitations that prohibit the ideal classroom design) and other challenges" (p. 17). These modifications may impact the effectiveness of the RBIS (Henderson & Dancy, 2009; Turpen & Finklestein, 2009).

At the same time, analysis within the physics education research community indicates that the students who have

\* Correspondence: [jchini@ucf.edu](mailto:jchini@ucf.edu)

Physics Department, University of Central Florida, 4111 Libra Drive, Orlando, FL 32816, USA

participated as research subjects in these studies are “better prepared mathematically and are less diverse than the overall physics student population” (Kanim & Cid, 2017, p. 1). Thus, curricula are typically developed and studied with one population of students and then used with another. While the literature supports active learning in general, how specific active learning strategies are actually implemented in contexts that may vary from where they were developed has not been well documented and their effectiveness in diverse settings has not been thoroughly demonstrated.

### Critical components of a research-based instructional strategy

In order to determine how an RBIS has been modified, one must first describe the essential elements of the RBIS. In fidelity of implementation work, such elements are called “critical components” (Century et al., 2010); in the concerns-based adoption model, such elements are called “components” of “innovation configurations” (Hall and Loucks, 1978). We will use “critical components” in alignment with recent STEM education literature (e.g., Stains & Vickrey, 2017; Borrego, Cutler, Prince, Henderson, & Froyd, 2013). Steps for identifying critical components can include reviewing the literature, user interviews, observations (Hall & Loucks, 1978), using an established list of critical components for the model, using expert opinion from publications, and using qualitative research methods (Mowbray et al., 2003 in Borrego et al., 2013).

For example, Borrego et al. (2013) used a literature review and panel of engineering and physics experts to propose 16 critical components for a broad set of 11 RBIS. They demonstrated that 13 of the 16 critical components discriminated between instructors who did or did not use the RBIS. Common critical components across several RBIS include “discuss a problem in pairs or small groups,” “work on problem sets or projects in pairs or small groups,” “participate in activities that engage them [students] with course content through reflection and/or interaction with their peers,” and “complete specially designed activities to ‘learn’ course concepts on their own without being explicitly told” (Borrego et al., 2013, p. 419).

### Propagating research-based instructional strategies

Developers should communicate the critical components of their RBIS to support adopters with their implementation. Stanford et al. (2017) state that developers should disseminate through both passive (e.g., journal articles and conference presentations) and active (e.g., workshops and sustained learning communities) pathways in order to promote adoption of the RBIS. However, developers often rely on mass-market communication channels and do not interact with potential adopters

(Khatri et al., 2016). Recent research has highlighted the role of incorporating instructor perspectives when developing RBIS and developing instructor support to promote propagation and sustained adoption (Khatri et al., 2016; Stanford et al., 2016). Specifically, Khatri et al. (2016) highlight that developers of well-propagated innovations (1) include collaborators and/or potential developers in creating and refining the RBIS, (2) engage in large-scale interactive and traditional dissemination, (3) recognize that instructors need support in order to successfully implement the RBIS, and (4) receive continual funding over an extended period of time. For example, a curriculum development team who recognized the informed nature of instructor decisions created a flexible curriculum with clear modification guidelines (Scherr & Elby, 2007).

Even if adopters are included in the development of an RBIS, the ways in which the RBIS is implemented in diverse settings are inevitably different; as Hutchinson and Huberman (1994) state, there is “no way to avoid the reconstruction of [a] practice as local staff make sense of it in their own context” (p. 34). This means that as instructors implement an RBIS into their course, they will inevitably modify and adapt the RBIS to fit the needs of their students and context. Thus, we need to hear from instructors about how they implement RBIS so that developers can be aware of and provide support for the modifications made during implementation.

### Examining fidelity of implementation

Stains and Vickrey (2017) recently argued for an increased focus on fidelity of implementation, the extent to which an evidence-based instructional practice is implemented as intended by the developer, to support making more valid claims about the effectiveness of these practices. They describe a framework and method for characterizing fidelity of implementation through identifying the critical components of the evidence-based instructional practice and measuring the extent to which individual implementations are faithful to those critical components.

### Exploring instructor’s purposeful modifications to RBIS

We agree with Stains and Vickrey (2017) that more nuance is needed to support claims about whether and why an evidence-based instructional practice supports student learning. However, the fidelity of implementation framework emphasizes the flow of information from the developer to the instructor about how the practice should be implemented. We argue that the reverse, the flow of information from instructors to the developers about the details of implementing interventions in diverse, real-world contexts, is equally important and not well captured by fidelity of implementation studies alone.

Our stance was inspired by our prior research on instructor's modifications of a popular RBIS called SCALE-UP (Student-Centered Active Learning Environment with Upside-Down Pedagogies), described in more detail below (Beichner et al., 2007). We found that many instructors modified the amount of time spent on lecture, student grouping methods, instructional materials (e.g., quizzes and content slides), and the use of teaching assistants (TAs) (Zamarripa Roman et al., 2017) compared to the model as described in the literature. However, the instructors were led to these changes due to their perceptions of student preparation, personal beliefs and values, and institutional facilitating conditions (i.e., funding allocations and scheduling conflicts).

### Purpose

In this commentary, we present a framework and method for exploring how and why instructors make purposeful modifications to their implementations of RBIS. We describe a framework for classifying modifications to RBIS and a method for analyzing instructors' reasoning related to the modification. We then demonstrate the use of the framework and analysis method on interview data about physics instructors' modifications to SCALE-UP. The knowledge generated from studies following this methodology will support curriculum developers and developers of RBIS in planning for variation in the contexts and student populations with which their intervention will be used, highlighting the "complexity and contextuality of learning" (Philip, Bang, & Jackson, 2018, p. 83).

### Instructional context: SCALE-UP

We draw on interviews with instructors teaching SCALE-UP introductory physics courses to demonstrate the usefulness of our framework and methodology. SCALE-UP is a "scaled up" version of studio-mode classes for larger enrollment courses (Beichner et al., 2007). Traditional introductory science courses feature separate lecture, laboratory, and recitation sections, and lecture time is often instructor focused. On the other hand, studio-mode courses typically combine some or all of the lecture, laboratory, and recitation activities into one meeting, allowing the instructor to flow between activities freely, and emphasize student-centered group work across these activities. Specifically, SCALE-UP courses are typified by (1) combined lecture, laboratory, and recitation time; (2) students spending a large portion of class time working together in groups; (3) minimized lecture time; (4) multiple instructors present during class time (including TAs) to monitor student learning; and (5) students presenting their work in class (Beichner et al., 2007). SCALE-UP has led to significant increases in student learning as measured by common physics concept

inventories (Beichner, 1999) and reduction in failure rates for women and students from underrepresented racial and ethnic groups (Beichner et al., 2007).

J.J.C. has taught SCALE-UP physics courses, participated in a research study about SCALE-UP instruction, and led a research study exploring other instructors' modifications to the SCALE-UP model. Her experiences and the developer's published literature on SCALE-UP informed our identification of the critical components, which are listed in [Appendix 1](#). More information about SCALE-UP can be found in Beichner et al. 2007, Beichner, 2008, or Knaub, Foote, Henderson, Dancy, & Beichner, 2016.

A recent study identified 314 departments at 189 institutions in 21 countries that described themselves as using or being influenced by SCALE-UP. While 30% of respondents described themselves as "users", about 38% identified as modified users, and 28% identified as "influenced by" SCALE-UP (Foote et al., 2014). Not surprisingly, over one third of the departments were in physics or related fields; however, the same study identified users from over a dozen disciplines. Foote et al. (2014) found about 35% of respondents reported learning about SCALE-UP from colleagues, 28% from a professional talk or workshop, 14% from the web, and 9% from the literature. This widespread dissemination and successful implementations (Neumeyer, Foote, Beichner, Dancy, & Henderson, 2014) make SCALE-UP an interesting case to explore for modifications of the RBIS elements.

### Identifying the *What* and *Why* in instructors' modifications of research-based instructional strategies

The purpose of this paper is to introduce a novel methodology to determine the modifications instructors make as they implement RBIS in their courses and the reasons that they make these modifications. We adapted a framework, which we will call the Modification Identification Framework (MIF; Stirman, Miller, Toder, & Calloway, 2013), to identify modifications instructors make. We also used Revealed Causal Mapping (RCM; Nelson, Nelson, & Armstrong, 2000) to create causal maps that help us understand why instructors made the identified modifications. In the following sections, we describe our adaptations to the MIF, our RCM process, and how the MIF and RCM can be used to explore instructors' experiences implementing RBIS.

### The Modification Identification Framework

Stirman et al. (2013) developed a framework for identifying modifications made to evidence-based interventions as they are implemented in new situations. The authors argue that as evidence-based interventions are

implemented in new situations with different target populations, personnel, contexts, and constraints, modifications are made to the intervention to make them relevant and useful in the new context. In their original article, Stirman et al. (2013) analyzed 32 articles published in the healthcare field and identified 258 modifications made during implementation of research-based interventions. Through an analysis of these modifications, Stirman et al. (2013) developed a coding scheme to classify the types of modifications. We propose that, after slight modifications are made to this framework, it can be used to identify changes made by instructors as they implement RBIS in their courses. Below is a description of the MIF for use in an educational setting (see Appendix 2 for a description of the modifications we made to Stirman et al.'s (2013) original framework).

### **Components of the MIF**

The MIF is composed of six main questions that allow the researcher to examine multiple facets of the modifications made to RBIS: (1) What is the point of reference? (2) By whom are the modifications made? (3) What is modified? (4) At what level of delivery are instructional practice modifications made? (5) What is the nature of the instructional practice modification? and (6) Which contexts are changed? (Stirman et al., 2013). (See Appendix 3 for a list of nuances of how these operationalizations were implemented in our sample study.) We identified changes from both the instructors' and the researchers' points of view. For example, if the instructor said they made a change, then this instance was coded with the MIF (instructor point of view). If the researcher identified that the instructor implemented a practice that was not in line with the RBIS critical components, then this instance was coded as well (researcher point of view). If the instructors' and researchers' views were at odds (e.g., if the instructor stated that they removed a practice that was aligned with the RBIS critical components but was not actually listed in the literature base as a critical component), then the researchers' point of view was followed.

We made two types of alterations to the original Stirman et al. (2013) framework. First, we changed the names of the categories to be more applicable to education. For example, in Stirman et al.'s (2013) framework, one of the levels of delivery is called "system"; we changed this level to "entire university or system/consortia of universities". In the healthcare setting, systems are typically the largest grain size at which changes to research-based interventions are made; similarly, in the educational setting, the entire university or system/consortia of universities is typically the largest grain size at which changes to RBIS are made. Second, we added

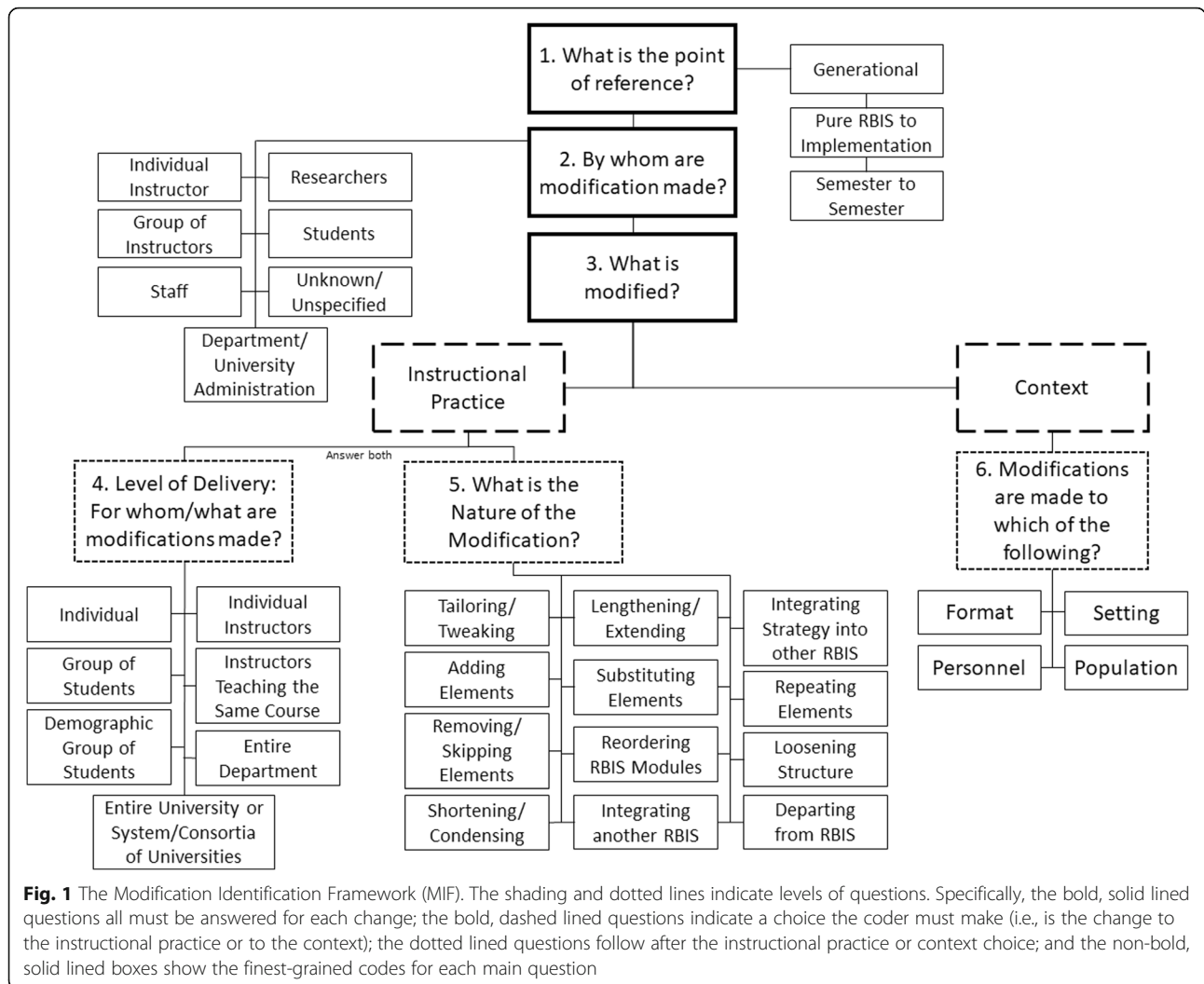
"Point of Reference" as an additional code. Figure 1 shows the MIF for use in an educational setting.

**Point of reference codes (Q1)** We added a category about point of reference to Stirman et al.'s (2013) MIF to identify the starting point from which the changes were made. This is a significant addition to the original framework in that it allows the researchers to identify what the modifications are relative to. Foote et al. (2014) found that the most common ways instructors learned about SCALE-UP were through their colleagues (corresponding to the Generational point of reference), a talk or workshop, the web, and literature (all corresponding to the Model point of reference). Adding the point of reference code allows researchers to see with respect to what reference point the change is being made and make claims about the relationship between dissemination and implementation. For example, some instructors make minor changes from semester to semester as they see how the RBIS works in their class (Semester to Semester point of reference), while others make changes that move their class away from the model RBIS (Model point of reference). Table 1 shows the point of reference codes with a description and an example. A single change can have multiple points of reference.

The Model and Generational points of reference do not correspond to a specific directionality for the change. For example, a change coded as a Model point of reference may correspond to a change that moves the implementation toward alignment with the RBIS or to a change that moves the implementation away from alignment. The Generational point of reference may correspond to a change that moves the implementation to more like or less like another instructor's implementation. The points of reference describe what the current implementation is being compared against. The Model point of reference compares the current implementation to the RBIS Model. The Semester to Semester point of reference compares the current implementation to a previous implementation of the RBIS by the same instructor. The Generational point of reference compares the current implementation to another instructors' implementation. The bidirectionality of these changes is an important addition because previous literature (i.e., Fidelity of Implementation) focuses solely on deviations from the "pure" RBIS model and does not discuss or identify why instructors move toward more RBIS model-aligned practices. This, in conjunction with the RCM, will allow researchers to make important claims about why instructors change the RBIS components that they change.

**Modification initiator codes (Q2)** The second question identifies which individual(s) is responsible for the modification. Stirman et al.'s (2013) original framework identified





a range of grain sizes of reformers; we translated these categories for the education setting as individual instructor, groups of instructors, staff, department and university administration, researchers, students, and unknown/unspecified. For the purposes of the sample study described below, we only focused on changes made by individual instructors; however, changes made by multiple groups can be identified using the MIF.

**Modification type codes (Q3–6)** The third question in the MIF identifies the aspect of the RBIS that is changed

as it is implemented in the new situation: the context or the instructional practice. *Context* changes (Q6) focus on the educational environment and are further categorized as impacting the setting, personnel, population, or format. Table 2 shows each context change code from Stirman et al.’s (2013) framework and the MIF, a description of the code, and examples including the point of reference codes. A single change could be described by multiple context change codes.

If the modification is an *instructional practice* change, there are two follow-up questions to classify the change.

**Table 1** Point of reference codes

Point of reference code	Description	Example
Generational	Modifications are made relative to the materials, delivery, and procedures passed down from other instructors.	An instructor modifying the lecture slides given to them by another instructor.
Model	Modifications are made relative to the main principles of the RBIS.	Lecturing during the entire SCALE-UP class period.
Semester to semester	Modifications are made relative to the implementation in a prior semester.	Modifying the methods of grouping students from one semester to the next.

**Table 2** Context change codes

Stirman framework	MIF context code	Description of code (MIF)	Example
Format	Format	The RBIS is implemented in a different set up or composition (e.g., changing the student to instructor ratio).	Using an RBIS with substantially more students than is suggested in the developers' model. (Model)
Setting	Setting	The RBIS is delivered in a different location or environment than originally intended.	Course is offered in a stadium-style classroom rather than in a SCALE-UP style room. (Model)
Personnel	Personnel	The RBIS is implemented by different instructors, staff, TAs, or other personnel.	Graduate TAs are replaced with undergraduate learning assistants. (Semester to semester)
Population	Population	The RBIS is used with a group of people with different demographics and characteristics than was originally intended.	The course is offered for an algebra-based course but the RBIS developer had only tested it in a calculus-based course. (Model)

The first (Q4) focuses on the level of delivery (e.g., for whom or what) at which the change is made; the possible levels of delivery are individual, groups of students, demographic group of students, individual instructors, instructors teaching the same course, entire department, and entire university or system/consortia of universities. Table 3 shows the level of delivery codes, a description of these codes, and an example including the point of reference codes.

The next follow-up (Q5) question for changes to the instructional practice focuses on the nature of the instructional practice modification, categorizing the methods by which the intervention can be modified. The categories are tailoring/tweaking, adding elements, removing/skipping elements, shortening/condensing, lengthening/extending, substituting elements, reordering RBIS modules, integrating another RBIS, integrating strategy into other RBIS, repeating elements, loosening structure, and departing from the RBIS. The term “element” in the MIF refers to salient or critical components of the RBIS. For example, in SCALE-UP, grouping

students is an element but the details of the grouping method are not an element. Table 4 shows these codes, a description of the codes, and an example including the point of reference code.

**Revealed Causal Mapping**

We employed Revealed Causal Mapping (RCM) to characterize the reasons why instructors make the changes they make to an RBIS as they implement it in their course. RCM is a method to uncover the mental maps that experts have about their subject domain. We chose RCM because the method values the knowledge and experience of experts, and in our study this translates to valuing the reasons instructors have for making changes to the RBIS as they implement it in their class. In contrast to the MIF, when using RCM, only the instructors' point of view is considered because the mental map that RCM will reveal should be the map of the instructor not the researcher. A goal of many agencies that fund RBIS development is to have successful strategies broadly implemented by instructors (Stanford et al., 2017);

**Table 3** Instructional practice changes—level of delivery codes

Stirman framework	MIF level of delivery code	Description (MIF)	Example
Individual	Individual	Modification is made for a single person.	Moving a female-identified student to a group with two male-identified students at her request when the model recommends having female students paired in groups. (Model)
Groups	Group of students	Modification is made for small groups within the classroom.	Regrouping specific students with group dynamics difficulties. (Model)
Clinics/units	Demographic group of students	Modification is made for a particular demographic, social group, or for people with a particular defining trait.	Adding course content and contexts relevant to the students' majors (e.g., biology context-rich problems) to materials inherited from another colleague. (Generational)
Individual practitioner	Individual instructors	Modification is made for all students in the instructor's courses.	An instructor using a different textbook than their colleague in all of their courses. (Generational)
Hospitals	Instructors teaching the same course	Modification is made for all instructors who teach a particular course at a particular time.	Changing all the sections of the same course to cover the same content with the same pacing. (Semester to semester)
Networks	Entire department	Modification is made for an entire department.	An entire department begins implementing an LA program for all of their courses. (Semester to semester)
Systems	Entire university or system/consortia of universities	Modification is made for an entire campus or group of universities.	Campus-wide initiative to start using clickers in class. (Semester to semester)

**Table 4** Instructional practice change—nature of modification codes

Stirman framework	MIF nature of modification code	Description (MIF)	Example
Tailoring/tweaking/refining	Tailoring/tweaking	Minor changes are made that leave all of the major principles intact while making the RBIS more appropriate, applicable, or acceptable in the particular classroom environment.	Changing the specific in-class problems for students to work in groups from a previous semester. (Semester to Semester)
Adding elements	Adding elements	Additional materials or activities are added that are consistent with the fundamentals of the RBIS.	Adding TAs to the course. (Model and semester to semester)
Removing/skipping elements	Removing/skipping elements	Main elements of the RBIS are not included in the implementation.	Removing TAs from the course. (Model and semester to semester)
Shortening/condensing (pacing/timing)	Shortening/condensing	Less time is dedicated to particular elements of the RBIS.	Spending less time lecturing during each class period. (Model)
Lengthening/extending (pacing/timing)	Lengthening/extending	More time is dedicated to particular elements of the RBIS.	Spending more time on group work during each class period. (Model)
Substituting elements	Substituting elements	An element of the RBIS is switched out for a different element.	Replacing an activity with physical equipment with a computational exercise. (Semester to semester)
Reordering of intervention modules or segments	Reordering RBIS modules	Modules/activities or concepts are used in a different order than what is recommended in the RBIS.	Rearranging the timing of course components (i.e., lecturing on a topic because students do an inquiry activity) from a schedule inherited from another instructor. (Generational)
Integrating another treatment into EBP	Integrating another RBIS	Implementing another RBIS into the RBIS of interest.	Integrating Process-Oriented Guided Inquiry Learning (POGIL; Moog & Spencer, 2008) into the SCALE-UP classroom. (Model)
Integrating the intervention into another framework	Integrating strategy into other RBIS	The RBIS of interest is implemented into a different RBIS.	Integrating SCALE-UP principles into a course that uses the Investigative Science Learning Environment (ISLE) model (Etkina & Van Heuvelen, 2004). (Semester to semester)
Repeating elements or modules	Repeating elements	Repeating a particular element of the RBIS.	Increasing TA preparation meetings from once per week to three times per week within a single semester as difficult course topics occur. (Model)
Loosening structure	Loosening structure	Features of the RBIS are not implemented as recommended in the RBIS but overall implementation maintains the main principles of the RBIS.	Changing the details of the grouping methods that are described in the RBIS model (e.g., allowing students to choose their own seats, only changing groups once per semester). (Model)
Departing from the intervention	Departing from RBIS	The RBIS is modified to the extent that the main principles of the RBIS are abandoned.	Teaching lecture and lab at separate times in different locations in contradiction to the RBIS model. (Model)

we see instructors as the experts on why they will or will not implement a particular strategy with a particular group of students. Therefore, it is crucial to explore their experiences implementing RBIS in diverse settings. In RCM, experts are interviewed about the topic of interest, their responses are examined for causal statements, and causal maps are developed from these causal statements.

#### **The Revealed Causal Mapping process**

Numerous articles describe the process for conducting RCM (see Allen, Armstrong, Riemenschneider, & Reid, 2006; Ghobadi & Ghobadi, 2015; Nelson, Nadkarni, Narayanan, & Ghods, 2000; Nelson, Nelson, & Armstrong, 2000). We combined the salient features of these four studies to produce our RCM process. Our combined RCM method is presented in Table 5.

The first step in RCM is to strategically select participants and collect data. Since RCM aims to create causal maps about the expert's topical domain, we must collect rich, qualitative data that will allow for this interpretation. Interviews and focus groups allow for this type of data collection. We collected the data for our sample study via interviews with instructors who were implementing SCALE-UP in their introductory physics courses. Next, if the amount of data is sufficiently large or if the data covers a broad range of topics, the data should be sampled. How the data should be sampled depends on the purposes of the project and the type of knowledge to be generated. However, typically, we want to include data from a broad range of participants to ensure that their associated broad views and causal maps will be included in the final causal map.

**Table 5** Our Revealed Causal Mapping (RCM) method

Process	Description	Operationalization in our sample study
1. Choose participants <sup>1,4</sup> and collect data <sup>1,3,4</sup>	Select participants who are experts in the area of interest for the study. Collect data in the form of interviews, focus groups, and/or artifacts.	We interviewed instructors with experience teaching SCALE-UP physics.
2. Purposively sample data <sup>4</sup>	Depending on the amount of data collected, sample data to include data from a broad group of participants.	We sampled data based on the instructors' institutional factors and SCALE-UP and overall teaching experience.
3. Identify causal statements <sup>1,2,3,4</sup>	Identify causal statements by searching for key linking words. Searching for: if, then, because/'cause, so, since think/thought, know/knew, use, believe, feel/felt.	Same
4. Separate causes and effects <sup>1,2,3,4</sup>	Separate causal statement into causes and effects by looking at directionality of linking words.	Same
5. Inter-rater reliability on identification of causal statements <sup>1,4</sup>	Investigate reliability of coding for causal statements.	First two authors independently coded each interview for causal statements and these responses were compared.
6. Identify relevant concepts <sup>1,3,4</sup>	Group frequently mentioned words from both causes and effects into categories (names are based on the participants' own words).	Same
7. Inter-rater agreement on concept identification <sup>1</sup>	Investigate reliability of concepts identified in causal statements.	Same
8. Construct raw causal map for each participant <sup>1,2,3,4</sup>	Use concepts and causal statements to construct causal map for each participant.	This was the end of our sample analysis as we only investigated four instructors. Future studies should continue the process.
9. Aggregate raw causal maps based on relevant sampling variables <sup>1,2,3,4</sup>	Combine causal maps for related participants to create an aggregate causal map. Participants can be grouped by variables of interest.	In future studies, we will group instructors by their institution and prior SCALE-UP and overall teaching experience.
10. Member checking <sup>2</sup>	Discuss causal map and aggregate causal maps with participants to check for appropriateness of interpretations and correctness of findings (Lincoln & Guba, 1985).	In future studies, we will follow-up with instructors to check our interpretation of their interviews and causal statements.
11. Analyze aggregated causal maps <sup>1,2,3,4</sup>	Use metrics such as the point of redundancy <sup>1,2,3,4</sup> , adjacency matrix <sup>2,3,4</sup> , reachability matrix <sup>1,2,3,4</sup> , centrality <sup>1,3</sup> , and density <sup>1,4</sup> to analyze and validate the aggregated causal maps.	In future studies, we will investigate the aggregated causal maps.

References: <sup>1</sup>Ghobadi & Ghobadi, 2015; <sup>2</sup>Nelson, Nelson, & Armstrong, 2000; <sup>3</sup>Allen et al., 2006; <sup>4</sup>Nelson, Nadkarni, et al., 2000

To make an RCM for each participant, their causal statements about the topic in question must first be identified. Causal statements include a cause of an event or phenomenon, an effect (i.e., the event or phenomenon), and a link between the cause and effect. Below is an excerpt from an interview with Instructor C:

We do estimations. Uh. Real-world problems I may -since they're most biology majors, I'll have problems that deal with certain aspects of biology. Where they have to use their knowledge of biology.

From this excerpt, we identified the following causal statement:

Since they're most biology majors, I'll have problems that deal with certain aspects of biology.

This means that the instructor added biology contexts to course problems (change) because most of the students enrolled in the course were biology majors (reason). Allen

et al. (2006) use keywords to identify causal statements by searching for the causal links. We used the following search terms: if, then, because/'cause, so, since, think/thought, know/knew, use, believe, feel/felt.

After each causal statement about the topic of interest is identified, these statements are broken into causes and effects. In terms of causality, the cause must come before the effect in time and the cause must lead to the effect. Continuing with the example from Instructor C, the cause is the fact that most students are biology majors, the effect is that the instructor added problems with biology contexts, and the linking word was "since". This causal statement identification and breakdown process should be conducted on the data from each participant included in the sample. To ensure that the identification and breakdown of the causal statements are reliable, an inter-rater reliability process should be conducted on this portion of the analysis. For example, multiple raters could conduct the causal statement identification and breakdown process and their findings could be compared to assess the reliability of their coding (Gwet, 2014).



Next, the causal statements for each participant are compared and grouped together to find larger-grained concepts that the participant talked about. This can be done by grouping frequently used words or ideas discussed by the participant. Continuing with our example from Instructor C, a concept could be “similar student demographics”. To ensure the reliability of the concept identification, an inter-rater agreement process should be conducted at this point. This inter-rater agreement process could take several forms, such as the process described above for the causal statement identification and breakdown or a collaborative discussion where expert researchers come to agreement on the categories through discussion.

For each participant, a raw RCM should be constructed next. The causal statements and concepts are used to construct a raw map for each participant by connecting each causal statement and concept together while keeping in mind the direction of causality of each statement. Figure 2 shows the raw RCM for the example from Instructor C:

Next, the raw RCMs are aggregated across participants to create a combined RCM. In a traditional RCM process, each participant’s raw RCM in the sample is included in the combined RCM (Nelson, Nelson, & Armstrong, 2000). However, there are some research projects and questions that would be better investigated by strategically grouping participants’ raw RCMs. For example, researchers may be interested to see if there are differences in how an RBIS is implemented at institutions with differing characteristics; thus, combining raw RCMs for instructors only at similar institutions would be valuable. As a method of triangulation to increase the validity, reliability, and rigor of the analysis and interpretation of the participants’ words and ideas, member checking should be conducted between the researchers and the participants (Lincoln & Guba, 1985). Member checking was not conducted for our pilot study because we did not have access to the instructors after the data were collected. We suggest that future studies be crafted such that researchers are permitted sustained access to the participants to allow for member checking.

The final step in the RCM is to analyze the combined RCM by calculating fit metrics, such as the point of redundancy (“the point at which further data collection would not provide additional concepts”; Armstrong, Riemenschneider, Allen, & Reid, 2007, p. 145), the adjacency matrix (a matrix that represents the frequency

of a connection between two concepts that is used as a measure of the strength of the casual statements; Nelson, Nelson, & Armstrong, 2000), the reachability matrix (“an indicator of the total strength of the connection between concepts”; Armstrong et al., 2007, p. 145), centrality (“a measure of the relative importance of a concept or how involved it is in the cognitive structure”; Armstrong et al., 2007, p. 145), and density (“calculated by dividing the number of links among constructs to the number of constructs in the map”; Ghobadi & Ghobadi, 2015, p. 334). We propose that the combination of the MIF and the RCM process will allow researchers to identify the changes instructors make to RBIS as they implement them in their courses and examine why instructors make these changes.

### Sample study

Below, we will present a sample study to demonstrate how the MIF and RCM can be used to investigate the changes that instructors make as they implement the SCALE-UP method in their introductory physics courses and their reasoning for making these changes.

To demonstrate the usefulness of the MIF and RCM, we implemented them on a small, yet diverse, subset of instructor interviews. To make sure the frameworks were useful in varied contexts, we purposefully selected four instructors’ interviews (from a larger sample of 43 instructors at nine institutions that self-identified as SCALE-UP users) based on their teaching experience and characteristics of their institutions. The interviews had been conducted as part of a larger study to explore introductory SCALE-UP physics courses in diverse contexts.

The interviews followed a semi-structured protocol to allow for spontaneous questions from the interviewer and to give participants the freedom to elaborate on their responses. Observations of the SCALE-UP style classes often occurred before the interviews were conducted which allowed us to investigate aspects of the instructors’ teaching that occurred in the class. The interviews also typically explored how the instructors began using SCALE-UP and what changes they have made as they have taught the course.

### Participants

We selected four instructors for this sample study from a larger dataset based on their teaching experience (i.e., number of years teaching physics and number of years



**Fig. 2** Instructor C’s raw Revealed Causal Map

teaching using SCALE-UP) and the characteristics of the institutions they teach at (i.e., residential campus status and student to instructor ratio in the class). We chose these four sampling criteria because the instructors' teaching experience will influence how they currently teach and implement SCALE-UP and the institutional factors will affect the constraints that are put on how the instructors can implement SCALE-UP in their classes (e.g., at a non-residential institution assigning group homework may be difficult as students do not have as much access to each other). First, we selected the four institutions by including one institution with each of the following characteristics: (1) residential with low student to instructor ratio, (2) residential with high student to instructor ratio, (3) non-residential with low student to instructor ratio, and (4) non-residential with high student to instructor ratio. Residential status was taken from the College Board website ("Create Your Road Map", 2018), and the categorizations of student to instructor ratios were determined based on the mean for our sample, where all undergraduate and graduate TAs, learning assistants (LAs), and faculty were counted as instructors. For example, if an institutions' student to instructor ratio was higher than the mean for the entire sample (which was 23 students per instructor), this institution was classified as having a high student to instructor ratio.

Next, we selected individual instructors to include in our sample by examining the instructors' overall physics teaching and SCALE-UP teaching experiences. We included one instructor from each of the following categories: (1) low physics and low SCALE-UP experience, (2) low physics and high SCALE-UP experience, (3) high physics and low SCALE-UP experience, and (4) high physics and high SCALE-UP experience. As with the student to instructor ratios, the high and low classifications were based on a comparison with the mean for our sample (which was 13 years overall teaching experience and 4 years SCALE-UP teaching experience). Through our sampling process, we selected the four instructors listed in Table 6.

### Coding and inter-rater reliability process

To identify the changes made by instructors during their implementation of SCALE-UP and their reasons for

these changes, two researchers (E.S. and B.Z.R.) implemented the MIF and RCM on the four instructors' interview data. The two researchers were a postdoctoral researcher (E.S.) and a physics graduate student (B.Z.R.), both with prior experience in qualitative research in physics education. Neither researcher conducted the interviews with the instructors. They both coded the same four instructors' interviews in their entirety.

We investigated the consistency or agreement between the two researchers' coding through an inter-rater reliability process to provide evidence for the reliability of our use of the MIF and RCM. We measured the inter-rater reliability of our MIF data with Gwet's AC1, a measure which is robust to low trait prevalence (i.e., when codes do not appear frequently in a sample; Gwet, 2002). Gwet's AC1 ranges from 0 to 1 (no agreement to perfect agreement, respectively), values of 0.61 to 0.8 indicate substantial agreement and values above 0.8 indicate near-perfect agreement (Gwet, 2014).

First, both researchers read about the MIF and RCM and their operationalizations in the literature base. Next, the two researchers trained in the coding with the MIF and RCM by coding two sample transcripts that are not presented in this article. After training, the researchers separately coded all four selected instructors' transcripts. For the MIF, they coded the entire transcript and generated a list of changes to the implementation of SCALE-UP discussed by each instructor.

From these changes, we focused on a small subset for analysis with the RCM method; specifically, we focused on changes related to the formation of student groups, TA and LA training, moving away from traditional lecture and toward SCALE-UP, and content changes to accommodate the needs and interests of the students. The unit of analysis for this coding was each individual question. For the RCM method, we coded interview questions where the selected changes were discussed; the unit of analysis for this coding was individual causal statement. The coders met after coding, discussed their independent coding, and came to agreement for the coding presented here.

### MIF reliability

Appendix 4 shows the results of the inter-rater reliability process for the MIF. The Gwet's AC1 values for all but

**Table 6** Instructor characteristics for sample study

Instructor <sup>a</sup>	Residential status	Student to instructor ratio	Physics teaching experience	SCALE-UP mode experience	Public/Private
A	Residential	10.6 (low)	1 year (low)	1 year (low)	Private
B	Residential	26 (high)	7 years (low)	4 years (high)	Public
C	Non-residential	17.8 (low)	30 years (high)	10 years (high)	Public
D	Non-residential	33 (high)	20 years (high)	2 years (low)	Public

<sup>a</sup>We refer to the instructors by generic identifiers to mask their identity and the identity of their institution. We chose not to use pseudonyms as we did not ask the instructors to generate pseudonyms, and we do not want to decide for the instructors what aspects of their identity to include in the pseudonym

one code were above 0.8 before discussion. These results show a substantial, and in most cases near-perfect, degree of reliability and provide evidence for the reliability of the coding with the MIF.

### **RCM reliability**

The RCM method reliability could not follow the same type of inter-rater reliability process as the MIF because we did not implement a priori codes. Instead, we compared the causal statements identified by the two researchers. After the two researchers had independently identified causal statements, they had identified 12 statements in common. In addition, one researcher (E.S.) identified six additional causal statements and the other researcher (B.Z.R.) identified one additional causal statement. After discussion, the researchers agreed on 17 causal statements. Finally, the researchers constructed the causal maps together and agreed on all of the maps presented in this paper.

### **Sample study findings**

Below is a brief discussion of our preliminary findings from the MIF and RCM coding.

#### **MIF findings: changes made by instructors**

We restricted our analysis to modifications made by individual instructors (*Q2: By whom are the modifications made?*). All of the changes identified were instructional practice changes (*Q3: What was modified?*). Since the current results are drawn from only four instructors, we cannot make generalizable claims. If we find that across the larger sample most changes are to instructional practice and not context, this result may be particular to RBIS that require a specialized space, like SCALE-UP (Knaub et al., 2016). Most changes were at the individual instructor level ( $41/45 = 91\%$ ; *Q4: For whom/what are the modifications made?*). Similarly, most of the points

of reference (*Q1*) were Model ( $25/45 = 55\%$ ) or semester to semester ( $18/45 = 40\%$ ). Table 7 shows the findings of the nature of modification code (*Q5*) for the four instructors' interviews.

Overall, the most commonly identified nature of modification code was loosening structure. Again, if this finding holds in the larger sample it may be true of RBIS like SCALE-UP that specify general instructional practices but not a curriculum. Most changes in this category were related to changing the method of forming student groups or changing/removing the methods for training TAs and LAs. (Examples for these codes are listed above in "The Modification Identification Framework" section.)

From all of the changes we identified with the MIF, we selected four types of changes to analyze with the RCM method because they were commonly identified or would be of interest to the discipline-based education research community: (1) forming groups, (2) TA/LA training, (3) changing the amount of lecture and group work to better align with the studio format, and (4) selecting topics of interest to enrolled students. We briefly describe our interpretation of the SCALE-UP model version for each below.

Related to forming groups, the literature suggests that instructors purposefully create groups of typically three students; that groups consist of a student from the top, middle, and bottom of the course; that students from underrepresented groups are not alone in their first group; and that groups are rearranged three to four times throughout the semester (Beichner et al., 2007). In our sample study, the forming groups' change was identified when instructors no longer followed the rules identified in the original developer's SCALE-UP literature and/or implemented their own rules (e.g., putting students who have the same dominant language in the same group).

**Table 7** Nature of modification findings

Nature of modification	Total	Instructor A	Instructor B	Instructor C	Instructor D
Tailoring/tweaking	11	3	3	4	1
Adding elements	2	1	0	1	0
Removing/skipping elements	3	0	0	2	1
Shortening/condensing	2	1	1	0	0
Lengthening/extending	3	1	0	1	1
Substituting elements	2	1	0	0	1
Reordering elements	1	1	0	0	0
Integrating another RBIS	4	1	2	1	0
Integrating strategy into other RBIS	0	0	0	0	0
Loosening structure	17	5	2	5	5
Departing from RBIS	0	0	0	0	0
Total	45	14	8	14	9

The SCALE-UP literature describes instructional teams where instructors are supported by TAs or LAs with whom they meet regularly to review content and pedagogy (Beichner et al., 2007). While TAs and LAs hold different positions in the classroom, for the purposes of our sample study, we will talk about the two together because they are both roles assisting the main instructor in teaching the class. Also, both TAs and LAs require training to be effective in the classroom, as described in the SCALE-UP model. We identified the TA/LA training change when the institutions did not have ongoing training for their TAs and LAs.

The SCALE-UP literature suggests limiting lecture to no more than 1 h per week (Beichner et al., 2007). We identified a change related to the amount of time spent on lecture in Instructor A’s transcript where the instructor talked about decreasing the amount of class time spent on lecture and increasing the amount of group work. This was a change moving toward the SCALE-UP model.

The SCALE-UP model suggests including problems relevant to real-world practices (Beichner et al., 2007). We identified this change when instructors modified their courses and course content for the students who enrolled. For example, some instructors talked about adding problems with biology contexts because of the presence of many biology majors in their courses. A description of this change was presented above in our description of the RCM process.

**RCM findings**

Below, we discuss the causal maps related to forming groups, TA/LA training, and lecture versus group work time. The findings of this sample study are preliminary since the causal maps need to be aggregated and the aggregate map(s) analyzed. Yet the preliminary findings illustrate how the MIF and RCM can be used in concert to guide developers in addressing the factors instructors consider when modifying their RBIS.

**Forming groups** Instructor A described changing the method of grouping students, as shown in the excerpt below:

Interviewer: So you form, you form groups randomly?

Instructor A: Yes with some idea not to have ... have the sexes mixed up and, and uh [Other Instructor] also likes to put majors together. If somebody’s in biochemistry, they’ll have a common interest. So he’ll like to match those up. Yeah I think that’s probably effective.

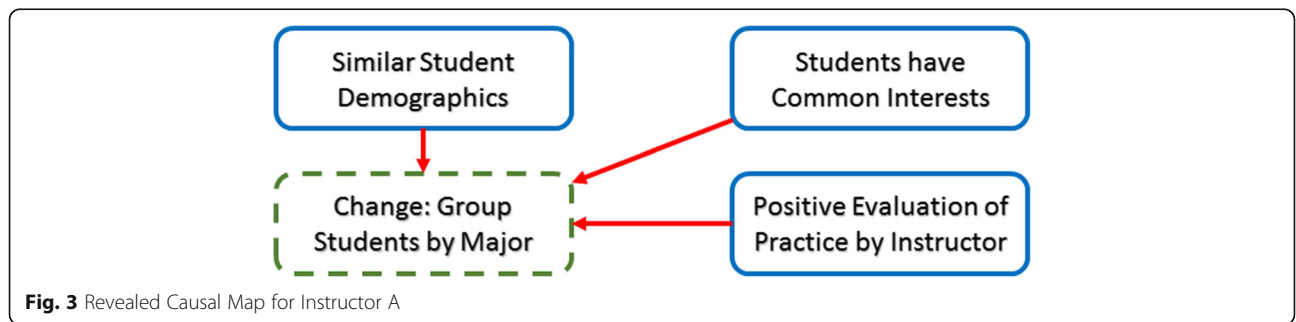
We combined the causal statements we identified (which are listed in Appendix 5) to make the causal map in Fig. 3.

From the first causal map, we can see that this instructor considers their students’ intended major as a proxy for common interests. In the instructor’s opinion, grouping by major can help students feel more comfortable due to common interests with their group members.

This causal map can alert curriculum developers that this instructor values grouping students with common interests. Viewed solely through a fidelity of implementation lens, this change diverges from the “pure” SCALE-UP model that has been validated and may affect the efficacy of the entire method. However, curriculum developers should be aware that instructors are making this change so that they can address it in their research (i.e., investigate the effectiveness of this practice) and in the instructor guide (i.e., give information for instructors about how to do this practice effectively or how instructors could meet their goal of forming groups with common interests through a method more in line with the suggested group formation practices).

**TA training** Instructor D described changing how they trained their TAs/LAs, as shown in the excerpt below:

Instructor D: But that’s the thing, and obviously I would like [TA], while he’s there, he does also do the wandering around and uh, and helping out um, I do not have, I do not meet with him nearly as often as I



**Fig. 3** Revealed Causal Map for Instructor A

meet with uh with my LA before, not right before every class, but twice a week, uh uh for twenty-thirty minutes each to go over what we are going to be talking about, I do not meet with, I meet with him once every other week or something like that. So he does not always know exactly what we are going to be doing, partly because I know he's only going to be there for an hour so it's not that big of a deal, but...

We combined the identified causal statements (listed in [Appendix 5](#)) to make the causal map shown in Fig. 4.

The second causal map indicates that this instructor is considering the amount of time TAs interact with students when weighing the consequences of cutting back the time spent on TA training. Instructors who make this choice may be considering the benefits of respecting the amount of time TAs are spending in class.

This map has important implications for how TAs are used in SCALE-UP classes. Specifically, the instructors' reasoning is predicated on the fact that the TAs at this institution are not given routine training in physics content or SCALE-UP pedagogy. How instructors handle this type of change is important information for the curriculum developers to know in order to investigate the effectiveness of these changes and/or to emphasize the truly salient components of SCALE-UP that lead to improved student outcomes. If TA training is found to be essential to the successful implementation of SCALE-UP, then curriculum developers could include predicted consequences of removing TA training so that instructors know the consequences when making their decisions.

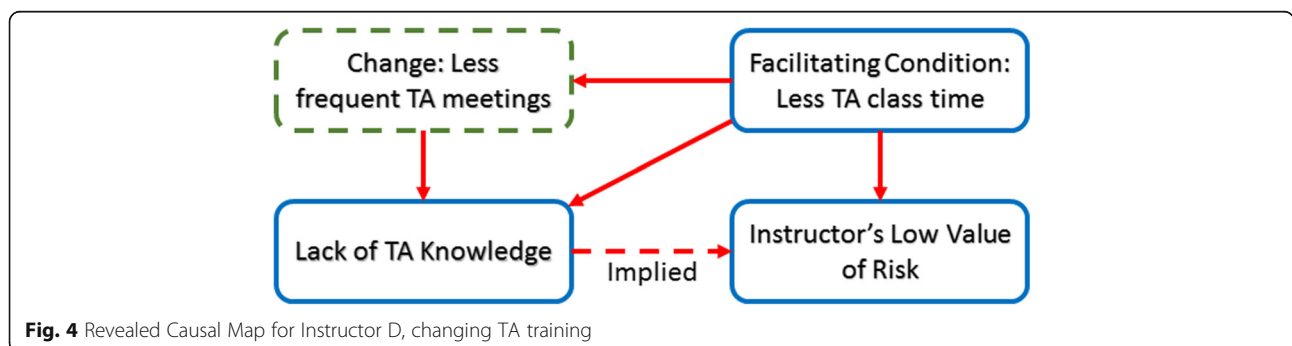
**Moving toward SCALE-UP** Instructor D described decreasing the amount of time spent in lecture to allow more time for group work, as shown in the excerpt below:

Instructor D: So I have my own notes and my own thinking and my own ideas before I start where the

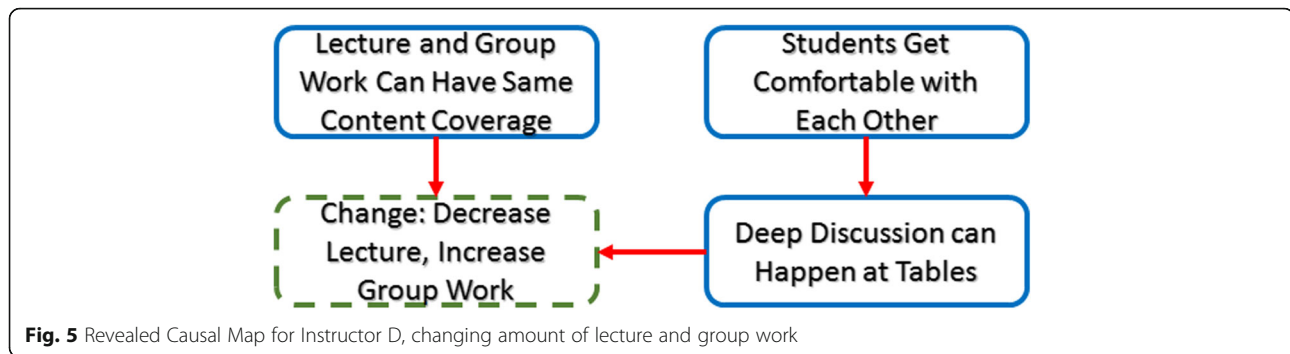
emphasis should be but I think um, that there is ... The first time I taught it there was there was I think more using the slides as a crutch a bit. The second time I taught it I learned that things are happening at those tables so it's not important to show, all the slides every time and spend time on each one [emphasis on 'all'], if if the same physics is covered by the discussion at the tables. And so I went a little bit in that direction. The uh, the thing I have to be careful of is that [short pause] sometimes I might see excellent discussions at some tables but not all tables. It's a little bit harder to make things uniform. But I think as the as the class goes on they they very quickly warm up to each other and they are at a table where they feel comfortable. And if somebody's missing because they are sick or something they really miss that person. So I think it's more getting in tuned to the importance of the table discussions. And I think the, the other thing I learned is that, try to balance a little bit, how much time I would spend at one table. And I think the first time I taught it I went in the direction 'Well I can't sit down at this table and spend ten minutes because there's eight other tables that won't get my attention.' But our staff is big enough so -so I think that I spent when I did really enter the discussion deeply at a table, I now spend more time there than I would have you know the first time I taught it [emphasis on 'really']. [pause] You know that -that's in part being comfortable with and trusting the other four people in the room, right to to be able to do what I can do at the tables.

The causal statements identified in this excerpt are listed in [Appendix 5](#).

We combined these statements to make the causal map shown in Fig. 5. Notice that some of the causal statements were grouped into the same concept, and thus, there are only three arrows in the causal map that correspond to five causal statements.







**Fig. 5** Revealed Causal Map for Instructor D, changing amount of lecture and group work

The third causal map demonstrates that this instructor considered content coverage when deciding to move away from lecture and toward student-centered discussion. This information can be used by curriculum developers, department chairs, and fellow faculty members as a way of tailoring their argument to their colleagues about the usefulness of reformed pedagogies. Specifically, Instructor D was convinced to move toward SCALE-UP (i.e., reducing lecture and increasing group work) by seeing that the same content can be covered in a lecture setting as in a group work setting.

### Limitations

One limitation of using the MIF and RCM method is that they must be implemented on RBIS that have been propagated. Khatri et al. (2017) describe three levels of specificity for RBIS: general (“movement or broad theoretical term in education literature with many possible implementations”), recognizable (“innovation has a name which is associated with a set of teaching practices, but has no central leadership”), and branded (“innovation name is associated with a set of teaching practices and has central leadership”; p. 5). In order for an RBIS to be analyzed in the method described in this paper, the RBIS must reach at least the recognizable level of specificity.

The RBIS to be investigated must be research-based in that there is literature supporting the strategy and there are descriptions of critical components available for the implementers to work from. The RBIS must also be implemented beyond the original developers for there to be changes from the original developers’ intention to identify with the MIF. The content and level of specificity of the guidelines for the RBIS should not affect the usability of the MIF and RCM method. However, if the literature about the RBIS is imprecise and contentious, this could lead to disagreements between researchers on the critical components which would affect the MIF findings. Khatri et al. (2017) found that most well-propagated RBIS require only pedagogical changes (26/43) and only one required changes only in content (1 in 43).

Another limitation is that using the MIF and RCM method does not prove which components of an RBIS are critical to effective implementation. The critical components can be determined a priori (as described in the section titled “Critical Components of a research-based instructional strategy”). Correlations between MIF findings and student-learning outcomes can be used to explore if the same components are necessary and sufficient in diverse contexts. Findings produced using the MIF and the RCM method do not connect to student-learning outcomes. However, these correlations can be made after using the MIF to identify changes instructors make while implementing an RBIS.

When creating a revealed causal map for a participant’s reasoning of why they made changes during the implementation of an RBIS, only changes that the participant identified can be studied. This is because the participant will not talk about a modification they do not know that they made unless made clear during the interview or focus group in which the data for analysis are collected. However, when analyzing data using the MIF, both participant- and researcher- identified changes may be analyzed.

### Conclusions

The combination of MIF and RCM can be used to highlight the voices of instructors who try an RBIS because they have been convinced it should support student learning, but run into challenges or unique opportunities in their local context. This is especially important since analysis suggests that, at least within the physics education research community, research has typically focused on students who are mathematically better prepared and less diverse than the general population of students who take physics courses (Kanim & Cid, 2017). We argue that failure to specifically elicit these instructors’ experiences risks “interest convergence”, “whereby the interests of non-dominant groups are only advanced in so far as they converge with the goals of the dominant groups” (Philip et al., 2018, p. 84).

We envision several types of studies that can be supported by this methodology. Used alone, the MIF provides

a systematic way to describe the changes instructors make to an RBIS. For example, we plan to use the MIF to describe the changes made by instructors who switched from using SCALE-UP with a calculus-based physics population to an algebra-based physics population. Used alone, the RCM method could highlight reasons instructors enter the modification process. We suggest that the MIF and RCM are best applied to questions of instructor modifications to RBIS in concert. Used together, the MIF and RCM can identify the instructor reasoning that leads to certain changes. For example, we plan to use the MIF to identify the types of changes instructors make to their use of lecture in SCALE-UP physics courses; when a change is identified, we will use RCM to describe the factors that led instructors to those changes. These results can be disaggregated based on factors of interest, such as typical mathematics preparation of students at the institution or proportion of students who live on campus, in order to develop recommendations for how instructors in similar institutional contexts may consider modifying the RBIS to support student success.

RBIS are typically developed and funded with the intention of broad implementation (Stanford et al., 2017). However, the uptake of RBIS has been slow (Stains et al., 2018). We suggest that the MIF and RCM can be used to identify experiences that lead instructors to thoughtfully and purposefully modify RBIS to work in their local context. Such studies can support RBIS developers in communicating factors that may warrant modifications to the RBIS and potentially successful modifications.

**Appendix 1**

**SCALE-UP critical components**

Below is a list of critical components for SCALE-UP as derived from the following literature: Beichner et al., 2007 and Foote et al., 2014.

1. Lecture, laboratory, and recitation time are combined and in the same room.
2. There is a special room that facilitates group interaction and collaboration.
3. Students work together in small groups.
4. Students acquire information outside of class and thus, instructors minimize the presentation of information in class.
5. Students practice applying knowledge in class.
6. Instructors (both faculty and TAs) monitor understanding during class.
7. Students solve real-world problems.
8. There is an emphasis on conceptual understanding.
9. Students are grouped together.
10. Students present their work during class time.

**Appendix 2**

**Comparison between Stirman et al. (2013) framework and the Modification Identification Framework (MIF)**

We made changes to some of Stirman et al.’s (2013) codes to make them appropriate for use in an educational context. We added the *point of reference* codes (Q1) to the Stirman et al. (2013) framework. The changes made to the *by whom are modifications made* codes (Q2) and *context change* codes (Q6) were to the names only; these changes were discussed in “The Modification Identification Framework” section. Tables 8 and 9 describe the changes we made to the *level of delivery* codes (Q4) and the *nature of modification* codes (Q5). The tables contain the Stirman et al. (2013) name for each code, the Modification Identification Framework (MIF) code, and a description of the modifications made to the Stirman et al. (2013) framework to produce the MIF.

We substituted “research-based instructional strategy” (RBIS) for “intervention” or evidence-based practice.

**Table 8** Instructional practice change—level of delivery

Stirman et al. (2013)	MIF	Description of adaptations from Stirman et al. (2013) to MIF
Individual patients	Individual	Individuals were changed from patients to a broader group of individuals including students, teaching assistants, and researchers.
Groups	Groups of students	Groups were specified as groups of students.
Clinics/units	Groups of students with similar demographics	Clinics and units were changed to be groups of students with similar demographics (e.g., ethnicity, gender, student major).
Individual practitioners	Individual instructors	Practitioners were modified to be instructors.
Hospitals	All or subset of faculty teaching the same course	Hospitals were changed to be all the faculty teaching different sections of the same course.
Networks	Entire department	Healthcare networks were changed to be the entire department in which the RBIS is implemented.
Systems	Entire university or system/consortia of universities	Healthcare systems were changed to the entire university or a system or consortia of universities.

**Table 9** Instructional practice changes—nature of modification

Stirman et al. (2013)	MIF	Description of adaptations from Stirman et al. (2013) to MIF
Tailoring/tweaking/refining	Tailoring/tweaking/refining	No modifications
Adding elements	Adding elements	No modifications
Removing/skipping elements	Removing/skipping elements	No modifications
Shortening/condensing (pacing/timing)	Shortening/condensing	Changed to be not limited to modifications of the amount of time.
Lengthening/extending (pacing/timing)	Lengthening/extending	Changed to be not limited to modifications of the amount of time.
Substituting elements	Substituting elements	No modifications
Reordering of intervention modules or segments	Reordering of RBIS modules or segments	No modifications
Integrating the intervention into another framework	Integrating strategy into other RBIS	No modifications
Integrating another treatment into EBP	Integrating another RBIS	No modifications
Repeating elements or modules	Repeating elements or modules	No modifications
Loosening structure	Loosening structure	No modifications
Departing from the intervention	Departing from the RBIS	No modifications

**Appendix 3**  
**The Modification Identification Framework (MIF)**  
**operationalizations nuances**  
*Level of delivery*

1. All changes made by the instructor for the entire class are at the Individual Instructor Level of delivery.
2. The Groups of students with similar demographics level changes are changes targeted and only applied for one demographic group (e.g., gender, ethnic, or social groups). When coding a Groups of students with similar demographics level of delivery change, the change must be implemented upon the specific population only instead of implementing the change on the entire class because of the presence of the specific population.

**Nature of modification**

1. Substituting elements is only coded if the element that is removed and the element that does the replacing are both explicitly described.
2. Shortening and Lengthening codes can refer to parts of the intervention and not necessarily the entire intervention because the intervention is the entire class.
3. For Integrating another RBIS, the RBIS being implemented into SCALE-UP must be research-based.

**Reference point**

1. For semester to semester changes, if we are given explicit information about the changes that

occurred over time (e.g., talking about incremental changes made on an exam over time) then we will code each change. If only implicit information is given, then do not code it.

**Other**

1. Miscellaneous changes category is for vague, undefined changes that are briefly mentioned by instructors.
2. Each coding represents one change and one change represents one instance of coding.
3. Only changes to SCALE-UP and not other approaches are coded.
4. Changes of the following form will *not* be coded:  
 Traditional lecture to studio model  
 Changes made at prior institutions

**Appendix 4**  
**MIF coding reliability data**

Appendix 4 shows the results of the inter-rater reliability process for the MIF. The Gwet’s AC1 values for all but one code were above 0.8 before discussion. These results show a substantial, and in most cases near-perfect, degree of reliability and provide evidence for the reliability of the coding with the MIF.

**Appendix 5**  
**Sample study causal statements**

Below are the causal statements we identified for the three examples from the sample study listed in the section titled RCM Findings.

**Table 10** Inter-rater reliability for the Modification Identification Framework (MIF): Gwet's AC1 values (per code)

Code			Gwet's AC1	
Instructional practice	Level of delivery	Individual	1	
		Group of students	1	
		Demographic group of students	0.97	
		Individual instructors	0.72	
		Instructors teaching the same course	0.97	
		Entire department	0.95	
		Entire university or system/consortia of universities	0.98	
		Nature of modification	Tailoring/tweaking	0.92
			Adding elements	0.96
			Removing/skipping elements	0.92
			Shortening/condensing	0.98
			Lengthening/extending	0.96
			Substituting elements	0.97
			Reordering RBIS modules	0.99
			Integrating another RBIS	0.99
			Integrating strategy into other RBIS	0.98
			Repeating elements	0.99
			Loosening structure	0.85
			Departing from RBIS	0.99
			Context	Format
Personnel	1			
Population	0.98			
Setting	0.99			

From Instructor A's excerpt in the section titled *Forming Groups*, we identified three causal statements:

Statement 1: If there are biochemistry students, then group them together.

Statement 2: Students will have things in common, so match them up.

Statement 3: I think grouping majors together is probably effective.

From Instructor D's expert in the section titled *TA Training*, we identified four causal statements:

Statement 1: TA does not know what will happen in class because he will only be in class for one hour.

Statement 2: Meet with TA every other week so the TA does not know what will happen in class.

Statement 3: TA will only be in class one hour so it's not a big deal.

Statement 4: TA will only be in class one hour so I meet with him every other week.

From Instructor D's excerpt in the section titled *Moving Toward SCALE-UP*, we identified five causal statements:

Statement 1: If some physics is covered in discussion then it's not important to show/discuss each slide.

Statement 2: If a group member is out sick then other group members really miss them.

Statement 3: Things happen at tables so not important to show/discuss each slide

Statement 4: Some physics is discussed at tables so went in that direction.

Statement 5: I think as class goes on students warm up and feel comfortable.

#### Abbreviations

ISLE: Investigative Science Learning Environment; LA: Learning assistant; MIF: Modification Identification Framework; POGIL: Process-Oriented Guided Inquiry Learning; RBIS: Research-based instructional strategy; RCM: Revealed Causal Mapping; SCALE-UP: Student-Centered Active Learning Environment with Upside-down Pedagogies; STEM: Science, technology, engineering, and mathematics; TA: Teaching assistant

#### Acknowledgements

We would like to thank Constance Doty for her help with this project.

#### Funding

This work is supported by NSF DUE- 1347515, 1347527 and 1347510.

#### Availability of data and materials

The dataset supporting the conclusions of this article is included within the article (and its additional files).

#### Authors' contributions

ES and BZR co-analyzed the presented data and co-authored the manuscript. EI assisted with the inter-rater reliability analysis and co-authored the reliability section. JJC assisted with the collection of the interview data and co-authored the manuscript. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

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

Received: 14 November 2018 Accepted: 18 March 2019

Published online: 11 April 2019

#### References

- Allen, M. W., Armstrong, D. J., Riemenschneider, C. K., & Reid, M. F. (2006). Making sense of the barriers women face in the information technology workforce: Standpoint theory, self-disclosure, and causal maps. *Sex Roles*, *54*, 831–844. <https://doi.org/10.1007/s11199-006-9049-4>.
- American Society for Engineering Education [ASEE]. (2012). *Innovation with impact: Creating a culture for scholarly and systematic innovation in engineering education*. Washington, DC: American Society for Engineering Education.
- Armstrong, D. J., Riemenschneider, C. K., Allen, M. W., & Reid, M. F. (2007). Advancement, voluntary turnover and women in IT: A cognitive study of work–family conflict. *Information and Management*, *44*. <https://doi.org/10.1016/j.im.2006.11.005>.

- Beichner, R. (1999). Student-centered activities for large-enrollment university physics (SCALE-UP). In *Proceedings of the Sigma Xi Forum: Reshaping undergraduate science and engineering education: Tools for better learning* Retrieved August 16, 2006, from <http://scaleup.ncsu.edu/>.
- Beichner, R. (2008). *The SCALE-UP Project: A student-centered active learning environment for undergraduate programs: An invited white paper for the National Academy of Sciences*.
- Beichner, R., Saul, J., Abbott, D. S., Morse, J. J., Deardorff, D. L., Allain, R., et al. (2007). Student-centered activities for large enrollment undergraduate programs (SCALE-UP). *Research-based Reform of University Physics*, 1, 2–39.
- Borrego M, Cutler S, Prince M, Henderson C, Froyd JE (2013). Fidelity of implementation of research-based instructional strategies (RBIS) in engineering science courses. *J Eng Educ* 102, 394–425. <https://doi.org/10.1002/jee.20020>.
- Century J, Rudnick M, Freeman C (2010). A framework for measuring fidelity of implementation: a foundation for shared language and accumulation of knowledge. *Am J Eval* 31, 199–218. <https://doi.org/10.1177/1098214010366173>.
- Create your road map. (2018). Retrieved from <https://bigfuture.collegeboard.org/>. Accessed 9 Nov 2018.
- Etkina, E., & Van Heuvelen, A. (2004). Investigative science learning environment. In *Forum on Education of the American Physical Society* (pp. 12–14).
- Foote, K., Neumeyer, X., Henderson, C., Dancy, M. H., & Beichner, R. (2014). Diffusion of research-based instructional strategies: The case of SCALE-UP. *International Journal of STEM Education*, 1(1), 1–18.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafo, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 8410–8415.
- Ghobadi, S., & Ghobadi, Z. (2015). How access gaps interact and shape digital divide: A cognitive investigation. *Behavior & Information Technology*, 34(4), 330–340.
- Gwet, K. L. (2002). *Statistical methods for inter-rater reliability assessment series 2*.
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (4th ed.). Gaithersburg: Advanced Analytics.
- Hall GE, Loucks SF (1978). *Innovation Configurations: Analyzing the Adaptations of Innovations*. Austin: University of Texas, Research and Development Center for Teacher Education.
- Henderson, C., & Dancy, M. (2009). The impact of physics education research on the teaching of introductory quantitative physics in the United States. *Physical Review Special Topics-Physics Education Research*, 5(2), 020107.
- Henderson, C., Dancy, M. H., & Niewiadomska-Bugaj, M. (2012). Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process? *Physical Review Special Topics-Physics Education Research*, 8(2), 020104.
- Hutchinson, J. R., & Huberman, M. (1994). Knowledge dissemination and use in science and mathematics education: A literature review. *Journal of Science Education and Technology*, 3(1), 27–47.
- Kanim, S., & Cid, X. C. (2017). The demographics of physics education research. Arxiv preprint. <https://arxiv.org/ftp/arxiv/papers/1710/1710.02598.pdf>.
- Khatri, R., Henderson, C., Cole, R., Froyd, J. E., Friedrichsen, D., & Stanford, C. (2016). Designing for sustained adoption: A model of developing educational innovations for successful propagation. *Physical Review Physics Education Research*, 12(1), 10112.
- Khatri, R., Henderson, C., Cole, R., Froyd, J. E., Friedrichsen, D., & Stanford, C. (2017). Characteristics of well-propagated teaching innovations in undergraduate STEM. *International Journal of STEM Education*, 4, 2.
- Knaub, A. V., Foote, K. T., Henderson, C., Dancy, M., & Beichner, R. J. (2016). Get a room: The role of classroom space in sustained implementation of studio style instruction. *International Journal of STEM Education*, 3, 8. <https://doi.org/10.1186/s40594-016-0042-3>.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park: Sage.
- Moog, R. S., & Spencer, J. N. (2008). *Process oriented guided inquiry learning (POGIL)*. Washington, DC: American Chemical Society; [New York]: Distributed by Oxford University Press.
- National Academies of Sciences, Engineering, and Medicine (NASEM). (2018). *Indicators for monitoring undergraduate STEM education*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/24943>.
- National Research Council. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington DC: The National Academies Press.
- Nelson, K. M., Nadkarni, S., Narayanan, V. K., & Ghods, M. (2000). Understanding software operations support expertise: A revealed causal mapping approach. *MIS Quarterly*, 24(3), 475–507.
- Nelson, K. M., Nelson, H. J., & Armstrong, D. (2000). Revealed causal mapping as an evocative method for information systems research. In *System sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on Systems Sciences*. Maui: IEEE.
- Neumeyer, X., Foote, K., Beichner, R., Dancy, M. H., & Henderson, C. (2014). Examining the diffusion of research-based instructional strategies using social network analysis: A case-study of SCALE-UP. In *ASEE Annual Conference and Exposition, Conference Proceedings*.
- Mowbray CT, Holter MC, Teague GB, Bybee D (2003). Fidelity criteria: development, measurement, and validation. *Am J Eval* 24, 315–340. <https://doi.org/10.1177/109821400302400303>.
- Philip, T. M., Bang, M., & Jackson, K. (2018). Articulating the ‘how’, the ‘for what’, the ‘for whom’, and the ‘with whom’ in concert: A call to broaden the benchmarks of our scholarship. *Cognition & Instruction*, 36(2), 1–6.
- President’s Council of Advisors on Science and Technology (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. In: Executive Office of the President (Ed.).
- Scherr, R. E., & Elby, A. (2007). Enabling informed adaptation of reformed instructional materials. In *Physics Education Research Conference Proceedings* (pp. 46–49).
- Stains, M., Harshman, J., Barker, M., Chasteen, S., Cole, R., DeChenne-Peters, S., et al. (2018). Anatomy of STEM teaching in North American universities. *Science*, 359(6383), 1468–1470.
- Stains, M., & Vickrey, T. (2017). Fidelity of implementation: An overlooked yet critical construct to establish effectiveness of evidence-based instructional practices. *CBE Life Sciences Education*, 16(1):rm1-11. <https://doi.org/10.1187/cbe.16-03-0113>
- Stanford, C., Cole, R. S., Froyd, J., Friedrichsen, D., Khatri, R., & Henderson, C. (2016). Supporting sustained adoption of education innovations: The designing for sustained adoption assessment instrument. *International Journal of STEM Education*, 3, 1.
- Stanford, C., Cole, R. S., Froyd, J., Henderson, C., Friedrichsen, D., & Khatri, R. (2017). Analysis of propagation plans in NSF-funded education development projects. *Journal of Science Education and Technology*, 26. <https://doi.org/10.1007/s10956-017-9689-x>.
- Stirman, S. W., Miller, C. J., Toder, K., & Calloway, A. (2013). Development of a framework and coding system for modifications and adaptations of evidence-based interventions. *Implementation Science*, 8(65). <https://doi.org/10.1186/1748-5908-8-65>.
- Turpen, C., & Finklestein, N. D. (2009). Not all interactive engagement is the same: Variations in physics professors’ implementation of peer instruction. *Physical Review Special Topics-Physics Education Research*, 5, 020101.
- Zamarripa Roman, B., Doty, C. M., Wilcox, M., Klinger, N., Pond, J. W. T., Von Korff, J. S., & Chini, J. J. (2017). Differences between the SCALE-UP model and instructors’ perceptions of implementation. In *Physics Education Research Conference Proceedings* (pp. 480–483).

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

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

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