RESEARCH   **Open Access**

Development of the Cooperative Adoption Factors Instrument to measure factors associated with instructional practice in the context of institutional change

Jacob D. McAlpin[1,2], John P. Ziker[3], John Skvoretz[2,4], Brian A. Couch[5], Brittnee Earl[6], Stephanie Feola[1,2], A. Kelly Lane[7], Karl Mertens[8], Luanna B. Prevost[9], Susan E. Shadle[6,10], Marilyne Stains[11] and Jennifer E. Lewis[1,2,4*]

Abstract

Background: Many institutional and departmentally focused change efforts have sought to improve teaching in STEM through the promotion of evidence-based instructional practices (EBIPs). Even with these efforts, EBIPs have not become the predominant mode of teaching in many STEM departments. To better understand institutional change efforts and the barriers to EBIP implementation, we developed the Cooperative Adoption Factors Instrument (CAFI) to probe faculty member characteristics beyond demographic attributes at the individual level. The CAFI probes multiple constructs related to institutional change including perceptions of the degree of mutual advantage of taking an action (strategic complements), trust and interconnectedness among colleagues (interdependence), and institutional attitudes toward teaching (climate).

Results: From data collected across five STEM fields at three large public research universities, we show that the CAFI has evidence of internal structure validity based on exploratory and confirmatory factor analysis. The scales have low correlations with each other and show significant variation among our sampled universities as demonstrated by ANOVA. We further demonstrate a relationship between the strategic complements and climate factors with EBIP adoption through use of a regression analysis. In addition to these factors, we also find that indegree, a measure of opinion leadership, correlates with EBIP adoption.

Conclusions: The CAFI uses the CACAO model of change to link the intended outcome of EBIP adoption with perception of EBIPs as mutually reinforcing (strategic complements), perception of faculty having their fates intertwined (interdependence), and perception of institutional readiness for change (climate). Our work has established that the CAFI is sensitive enough to pick up on differences between three relatively similar institutions and captures significant relationships with EBIP adoption. Our results suggest that the CAFI is likely to be a suitable tool to probe institutional change efforts, both for change agents who wish to characterize the local conditions on their respective campuses to support effective planning for a change initiative and for researchers who seek to follow the progression of a change initiative. While these initial findings are very promising, we also recommend that CAFI be administered in different types of institutions to examine the degree to which the observed relationships hold true across contexts.

---

*Correspondence: jennifer@usf.edu

[1] Chemistry, University of South Florida, 4202 E Fowler Avenue CHE201, Tampa, FL 33620, USA
Full list of author information is available at the end of the article

Introduction

Evidence-based instructional practices (EBIPs) are techniques that have been reported in the literature to consistently improve students' academic performance and affective outcomes in a wide variety of STEM disciplines (Freeman et al., 2014; Schroeder et al., 2007; Stains et al., 2018; Theobald et al., 2020). While EBIPs show promise in attracting and retaining more students in STEM, they have yet to become the predominant mode of instruction in college STEM courses (Borrego et al., 2010; Durham et al., 2017; Henderson & Dancy, 2009; López et al., 2022; Stains et al., 2018). Within the literature on institutional change efforts designed to promote the use of EBIPs in STEM education, several researchers have noted that discussions about teaching among faculty members impact practice (Kezar, 2014; Sachmpazidi et al., 2021). Some of these studies have examined departments within a university to map out patterns of social interactions to characterize how faculty talk about teaching among each other (Andrews et al., 2016; Knaub et al., 2018; Ma et al., 2018; McConnell et al., 2019; Mestre et al., 2019; Middleton et al., 2022; Quardokus & Henderson, 2015). Others have considered wider patterns of influence in teaching practice by looking at different universities across a discipline (Hayward & Laursen, 2018) or examining more specialized national initiatives such as The POGIL Project (Shadle et al., 2018). Members from our team have reported evidence that faculty members are influenced in their teaching practice by their discussion partners (Lane et al., 2019) and that faculty members who use EBIPs are more likely to seek conversations with other faculty who use EBIPs (Lane et al., 2020).

It is our goal to supplement this investigation of how faculty member's discussion of teaching affects their teaching practice with aspects of how individual faculty members perceive aspects of their local environment and institutional structures. For the theoretical basis of our research, we use the CACAO (Change–Adopter–Change Agents–Organization) model by Dormant (2011) to describe three constructs of interest to institutional change along with a description of adoption of a new behavior. The constructs include conceptualizations of the self-perception of mutually beneficial actions (Jackson & Zenou, 2015), the interconnectedness of success (Aktipis et al., 2018) and a faculty member's perception of institutional readiness for change (Landrum et al., 2017). As individual faculty members' perceptions of these constructs will likely be affected by their peers, we will consider these constructs in the context of investigating the social connections and networks of the faculty. As our goal is to better understand how faculty approach instructional practice in undergraduate STEM, we desired to characterize a large sample of faculty over several institutions according to these constructs. To accomplish this task, we developed the Cooperative Adoption Factors Instrument (CAFI), a tool that can efficiently and simultaneously measure each of these constructs. This report details the steps in the development, internal structure analysis, and interpretation of scores for our instrument.

Theoretical framework

Institutional change

In order to characterize adoption of EBIPs, we start by considering the CACAO model by Dormant (2011). The CACAO model combines several aspects of the change literature, including Roger's diffusion of innovations (Rogers, 2003) and the purposeful implementation of Kotter (Kotter & Cohen, 2002), to provide a useful framework for those designing change initiatives (Shadle et al., 2017; Viskupic et al., 2022). The model also emphasizes the importance of considering change at several levels as suggested in the name of the model itself: **C**hange (the new system, process, etc., that is desired), **A**dopters (the ones who will be implementing the change), **C**hange **A**gents (the ones promoting the change), and **O**rganization (the institution in which the others exist).

When considering the adoption of an innovation, the CACAO model describes a series of discrete steps taken by individuals: (1) awareness, (2) curiosity, (3) mental tryout, (4) hands-on tryout, and (5) adoption (Dormant & Lee, 2011). This characterization draws from several sources including Diffusion of Innovations (Gibbons et al., 2017; Rogers, 2003), the Concerns-Based-Adoption-Model (Anderson, 2014; Hall & Hord, 1987; Hord et al., 1987), and Lewin's 3-stage model of change (Lewin, 1947). Landrum et al. (2017) has previously used the CACAO characterization of steps to develop a Guttman scale of an individual faculty member's level of EBIP adoption in post-secondary education. This Guttman scale uses self-report from faculty on their degree of knowledge and use of EBIPs in their classrooms through a series of items matching an expected progression as informed by the steps of adoption in the CACAO model. In the development of the CAFI, we used this scale in order to characterize EBIP adoption among faculty.

The CACAO model also emphasizes understanding eventual adopters of change and enabling change

agents within their institutional context. Within this framework, the ideal organizational change is characterized by "relative advantage", positive "social impact", "compatibility", "simplicity", and "adaptability", all from the perspective of the adopter (Dormant & Lee, 2011) as illustrated in Fig. 1. In this scheme, relative advantage refers to the benefits of using the new method over the old way of doing things. If a change offers little benefit over the status quo, it is unlikely to see wide adoption. Social impact describes how implementing a change can affect the social relations of the potential adopter. A change which may result in harm to an adopter's social relationships is less likely to be implemented than a change with a perceived neutral or positive effect. Compatibility refers to how well a change fits into current practices, with more compatible changes being more easily adopted. Simplicity is a direct characteristic of the change. Complex changes are less likely to be implemented than simpler ones which potential adopters can more easily comprehend. Finally, adaptability is another direct characteristic of the change, referring to the degree to which adopters may customize it to the particular context. Here, changes which are adaptable are more likely to be implemented than more rigidly structured ones.

Using this framework, we seek to analyze STEM instructional reform efforts by characterizing relative advantage, social impact, and compatibility. In the interest of increasing the utility of the CAFI in a variety of contexts, we did not incorporate simplicity and adaptability, as these features are strongly dependent on the particular EBIP reform being implemented. For relative advantage, social impact, and compatibility, we have operationalized a particular element of each as a target of measurement, namely strategic complements, interdependence, and campus climate, respectively. These operationalizations are conceptualized in a way that acknowledges the embeddedness of faculty members in a social network.

Strategic complements
The concept of strategic complementarity describes how a faculty member views the relative advantage of taking on an action or behavior based on the actions or behaviors of other faculty around them. This terminology was originally coined by Bulow et al. (1985) and expanded into social networks by Jackson and Zenou (2015). The underlying foundation for strategic complements is within game theory, a theoretical framework developed in the mid-twentieth century (von Neumann & Morgenstern, 2007), and this concept has been applied to problems ranging from environmental dilemmas (Hardin, 1968) to Cold War nuclear deterrence (Schelling, 1966) and food sharing among traditional hunter-gatherers (Hames & McCabe, 2007; Ziker & Schnegg, 2005). In our context, we will consider how strategic complements apply to faculty determining the relative advantage of applying EBIPs in their practice. When EBIP adoption by others is a strategic complement, adoption has a relative advantage over not adopting. This relative advantage comes from the perception that use of EBIPs has greater benefits and/or reduced cost than if fewer faculty were using them.

Interdependence
If adoption of a new behavior may affect the social relationships of the adopter, the potential adopter will consider the positive or negative impact before adopting the new behavior. The CACAO model refers to this consideration as the social impact of change (Dormant & Lee, 2011). Adoption of EBIPs by faculty requires they first know about EBIPs. This knowledge can come from the exchange and sharing of ideas and experience between change agents and potential adopters of EBIPs. López et al. (2022) stress that network expansion is an important driver for the change process. When individual faculty members expand their social networks, they have the potential to gain access to resources, including knowledge of and support for using EBIPs. However, the sharing of ideas and experience among faculty members is likely to depend on the "comfort level" faculty have about sharing their teaching experiences and whether they feel they have some influence on the outcomes of others. When faculty do not have this "comfort level" with their peers, adoption of EBIPs
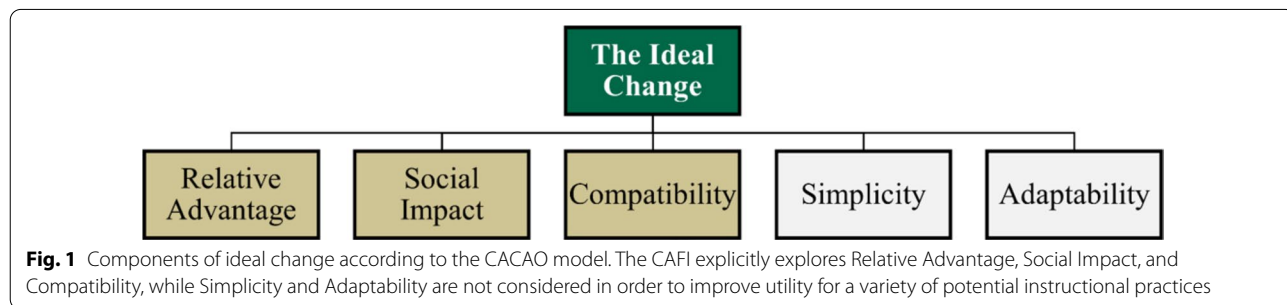
Fig. 1 Components of ideal change according to the CACAO model. The CAFI explicitly explores Relative Advantage, Social Impact, and Compatibility, while Simplicity and Adaptability are not considered in order to improve utility for a variety of potential instructional practices

or conversations about implementing EBIPs could result in a negative social impact. The resulting silence among peers interrupts the process of institutional change. The tendency of faculty to make trust-based decisions is supported by others in the institutional change literature (e.g., Kezar, 2014), and we identify trust as another key change-readiness metric relevant to planning successful higher education change initiatives. However, directly measuring trust is difficult, and widely used trust scales have garnered criticism (Naef & Schupp, 2009). To characterize this aspect of faculty relationships, we focus on interdependence, the idea that people see each other's success or failure as intertwined, which is commonly used in the majority of definitions of interpersonal trust (Boon & Holmes, 1991). We adapted a scale of interdependence (Aktipis et al., 2018) that was designed to be modified for use in a variety of settings. Using this scale of interdependence, Aktipis et al. (2018) found that interdependence can predict helping behaviors more strongly than other common measures, such as kinship.

Climate

Another consideration for faculty adopting EBIPs is whether a faculty member perceives EBIPs as compatible with their current classroom practice and/or with the climate regarding teaching on the campus as a whole (Walter et al., 2021). A faculty member's perception of the campus climate toward teaching is shaped by the faculty member's interactions with peers, and the interactions can have an impact on the implementation of innovative practices (McConnell et al., 2019). If a faculty member perceives the campus climate as supportive of teaching, then it is possible the faculty member will invest more time and energy into teaching and potentially adopting EBIPs (Sturtevant & Wheeler, 2019). A positive campus climate might also suggest that greater experimentation with teaching methods would be allowed and rewarded, while a negative climate might deter experimentation. Landrum et al. (2017) developed an instructional climate survey with the understanding that both climate and personal change characteristics can be helpful to campus change agents in assessing the current STEM landscape of faculty practices. Through this project, we seek to investigate how a faculty member's perception of climate in our contexts is related to the use of EBIPs.

Research questions

Each of the constructs discussed (strategic complements, interdependence, and campus climate) can have an impact on willingness to adopt a reformed teaching practice. Therefore, a goal of this research is to investigate the connection between these constructs and the measure of EBIP adoption. This study aims to provide an initial insight into the interplay between these constructs and EBIP adoption among faculty teaching undergraduate STEM. Additionally, as social connections can impact teaching practices (Kezar, 2014; Lane et al., 2019, 2020), we also want to consider how network measures predict EBIP adoption. As previous work has shown that STEM faculty members who have more extensive social connections tend to teach in more learner-centered ways than their peers (Middleton et al., 2022), understanding how faculty members are connected to each other is important to understand the diffusion of EBIP adoption. If the faculty who use EBIPs are serving as opinion leaders in their departments, this is promising for the future spread of EBIPs in the departments. While formal positions like department chair can serve as proxy for identifying opinion leaders, we want to better understand with whom faculty are having conversations outside the formal structure of their departments. Therefore, we included in our survey a prompt asking the respondent with whom they interacted about teaching matters. We then used the amount that a faculty member was identified as a target for teaching discussion as a measure of opinion leadership.

Discussion of theory related to institutional change in higher education generally falls into one of two categories, namely "change theory" or "theory of change" (Reinholz & Andrews, 2020). A change theory applies understanding from several theoretical and empirical reports over time to describe the process of change beyond a particular project or location. In contrast, a theory of change is generally project-specific and is focused on how to effect change within a specific context. It is our intention that our instrument, grounded in the change theories that are incorporated into the CACAO model, can serve as a starting point for change agents to build a theory of change within their institution. The constructs measured by our instrument all link to ideas around institutional change, and this instrument is intended to allow change agents to monitor a change initiative as it progresses at their institution.

The implementation of an instrument allows researchers to quickly and easily characterize a sample based on many constructs of interest. To optimize response rates, we prioritized making the instrument relatively short while incorporating multiple scales. As part of the development process, it is important to consider aspects of validity before interpreting scores (AERA, 2014; Arjoon et al., 2013) Additionally, this report seeks to provide some baseline levels for these constructs that future researchers can use as a reference point when

interpreting their own data. Here, we detail the steps we took to provide evidence for content and internal structure validity for the instrument along with some preliminary observations based on the observed outcomes. Specifically, we address:

1. To what extent are we able to provide evidence of content validity and internal structure validity for our instrument simultaneously measuring perception of strategic complements, interdependence, campus climate, and EBIP adoption?
2. Do the factor scores cover a range of values in their construct, enabling characterization of a range of faculty perceptions?
3. How do these factor scores and opinion leadership relate to EBIP adoption for this sample?

Methods
Instrument development and administration
Sampling frame
The sampling frame for our study consists of five STEM disciplines (biological sciences, chemistry, earth sciences, mathematics, and physics) at three large public research universities. The universities will be referred to as Uni1, Uni2, and Uni3. Each university is in a different region of the United States and has a Carnegie classification of high or very high research activity. Undergraduate enrollment ranges from around 20,000 to around 40,000. Institutional change initiatives designed to increase adoption of EBIPs were ongoing at each university at the time of data collection. All faculty with a teaching role in the semester of administration were considered eligible participants. Institutional Review Board approvals from each institution were obtained to conduct the research.

Instrument development
We piloted an initial version of the CAFI at Uni1 in October 2017. The pilot instrument included scales for strategic complements, interdependence, and EBIP adoption but not climate. We also included items addressing EBIP knowledge and use from a previous report on social influence (Lane et al., 2019) on this pilot version. The sample for the pilot was composed of faculty in departments that were not targets for the final version of the CAFI. Eligible participants received email invitations to respond to the pilot instrument administered via Qualtrics. This pilot administration received results from 154 respondents with 148 complete responses.

The first set of items in this pilot was designed to probe faculty members' perceptions of types of strategic interactions based on game theory. We originally constructed a series of 14 seven-point Likert-scale items in the fall of 2017 describing a variety of possible strategic interactions between faculty over EBIPs, including strategic complements, strategic supplements, and competitive exclusion (Jackson & Zenou, 2015). From the received responses, we conducted iterated factor analyses and reliability analysis using SPSS 24 on the 14 items. We examined all items together, each subscale separately, and then tested combinations of items from the different subscales. This analysis showed that items aimed at strategic complements and strategic supplements loaded onto two factors, but those two factors crosscut the groups of items. We also found that items originally aimed at a competitive exclusion subscale loaded onto one factor but had low reliability ($\alpha < 0.5$, Tavakol & Dennick, 2011). For the final version of the CAFI, we settled on four items originally tagged as strategic complements, one item originally tagged as a strategic supplement, and one item focused on the value of EBIPs for student learning into a scale that loaded onto one factor and had high reliability ($0.7 < \alpha < 0.9$, Tavakol & Dennick, 2011).

The next set of items was designed to probe perceptions of interdependence, a component of trust. These items were developed as part of the Human Generosity Project (Aktipis et al., 2018) and have been shown to predict helping behavior better than kinship or reciprocity. For these items, "your closest department colleague" was used as the person of reference. These items are scored on a 7-point Likert scale. Analysis of data collected in this pilot suggested no adjustments were necessary for these items.

In addition to collecting data on EBIP adoption through a Guttman style scale (Landrum et al., 2017), we supplemented the pilot of the CAFI with another set of items previously used to characterize EBIP knowledge and EBIP use separately (Lane et al., 2019). When analyzing the EBIP adoption scale relative to the separate EBIP knowledge and EBIP use scales, we see a correlation of 0.608 for EBIP adoption and EBIP knowledge and a correlation of 0.708 for EBIP adoption and EBIP use. These values are large enough to suggest we are characterizing similar constructs but not so strong to suggest redundancy.

All items and analyzed data from the pilot were then reviewed by an external advisory board composed of experts with respect to institutional change within higher education or social networks in educational contexts. Based on feedback from this board, items to address the campus climate were added. These items were added to help us consider the influence of the faculty member's perception of the compatibility between EBIP adoption and the faculty member's perception of the general campus climate for teaching. This section included six items to characterize the perception of campus climate toward

teaching. These items were chosen from analysis of data collected by Landrum et al ([2017](#)) by performing a factor analysis and reliability analysis of the original data from Landrum et al. We chose the items that contributed most strongly to one factor and contributed to the strongest reliability score. The items for this scale were answered on a 7-point semantic differential ranking perception between two opposing statements.

Before finalization of the CAFI, a section to collect data on faculty members' social networks was also added. For network data, each respondent was asked to identify individuals that satisfied the following prompt: "During the most recent academic year, I discussed instructional activities (e.g., teaching strategies, student learning, grading, student achievement) with the following colleagues." This prompt was provided for colleagues fitting three different descriptions: within department, outside department but within university, and outside university. To identify faculty members within department, respondents were provided a list of faculty members within their department whom they could check off as satisfying the prompt. For the other categories, respondents were allowed up to seven open responses for each as the nature of these categories do not allow for a prepopulated list. Relatively few respondents filled in all seven open response opportunities suggesting that limiting responses to seven did not meaningfully limit our data.

The final version of the CAFI as seen by participants is included as Additional file [1](#). During the Spring 2018 semester, 488 faculty were invited by email to participate by completing the CAFI administered via Qualtrics. A representative from our research team attended a departmental meeting in most of the sampled departments to encourage participation. As an incentive for completion, faculty at Uni1 received a $10 remuneration on the campus identification card for completing the survey in addition to a lunch being provided for all departments which reached 80% response or higher. At Uni2, faculty who completed the survey were given a $20 gift card for Target stores. Faculty at Uni3 received no compensation for completing the instrument. These differences reflect IRB interpretations of the allowability of incentives.

A total of 296 faculty gave complete or partial responses to the instrument. The response rate by campus ranged from 75% for Uni1, 62% for Uni2, and 50% for Uni3. From each of the three universities, we received a similar number of responses, meaning that each contributed about a third of the data (33%, 33%, and 34%, respectively). Breaking down our total responses by academic discipline, we saw fairly even representation from biology (22%), chemistry (25%), geology (20%), and mathematics (22%) with a smaller percentage coming from physics (12%); this suggests the data represent perspectives from

a range of STEM disciplines (numbers add to greater than 100% due to rounding). The sample had more male (68%) than female (32%) respondents which is reflective of the demographics of the eligible faculty being 72% male. Additionally, we received responses from a mixture of tenured/tenure-track (75%) and other faculty ranks (25%). The median time to complete the survey across all participants was just under 12 min.

The items used in our instrument represent both newly developed items and sets of items that have not previously been explored together. Therefore, we chose to use an exploratory factor analysis (EFA) to investigate how to explain the observed variance parsimoniously followed by confirmatory factor analysis to further support the internal structure (Brown, [2015](#)). Based on simulation studies, a three-factor instrument with six indicators per factor requires between 100 and 190 respondents to achieve appropriate statistical power with the range representing high to low factor loadings (Wolf et al., [2013](#)). Therefore, we expect analyzing around 150 responses will provide appropriate statistical power for each factor analysis. From the received 296 responses, we randomly split them into two halves of 148 responses. After completing the EFA on the first set of data, we performed confirmatory factor analysis (CFA) on the reserved portion to see if the data matches a prespecified model.

Statistical analysis

For analysis, responses to all scale items were coded on a 1–7 scale with 4 representing a neutral response, and negatively phrased items were reverse coded for ease of interpretation. Both EFA and CFA were run in Mplus 7 (Muthén & Muthén, [2012](#)). In both EFA and CFA, a full-information maximum likelihood (FIML) estimator was used as it is robust in estimating factor loadings for most analyses and can be used in cases with missing data without requiring listwise deletion. For EFA, a Geomin oblique rotation was used to improve interpretability while considering potential correlation among factors. While there is no absolute cutoff in loading for determining a factor, we chose a loading of 0.32 to indicate that a particular item loaded into a factor (Brown, [2015](#)). After results were obtained for the EFA, a CFA was performed on the reserved data. Details about the steps taken to address model fit are addressed in the results section.

One way to support results from a factor analysis is measures of internal consistency. While Cronbach's alpha (Cronbach, [1951](#)) is popular and included in this report, it is calculated under the assumption of an essentially tau-equivalent model which is rarely observed (Komperda et al., [2018](#)). Therefore, we also report values for omega (also referred to in literature as McDonald's omega or hierarchical omega (McDonald, [2013](#))) and

Raykov's rho (also seen in literature as Bollen or Raykov's omega (Bollen, 1980; Raykov, 2001). These supplemental measures have a similar interpretation with values closer to one being good and no absolute cutoff available.

Items to measure EBIP adoption were adapted from previous work in the development of a Guttman style scale (Landrum et al., 2017) and were analyzed separately from the other items. On this EBIP adoption scale, the respondent was presented with a series of items where agreement on later items suggests agreement with previous statements (e.g., use of EBIPs implies knowledge of EBIPs). Deviations from the pattern, where agreement on later items is accompanied by disagreement on earlier items, would indicate a lack of reliability in the scale. The number of 'yes' responses on this scale were added together to produce an EBIP adoption score. In characterizing the quality of results from a Guttman scale, recommended statistics include the coefficient of reproducibility (CR), minimal marginal reproducibility (MMR), percent improvement (PI), and coefficient of scalability (CS) (McIver & Carmines, 2011). These statistics would flag deviations from the pattern and can be used to support that a particular Guttman scale is functioning as a unidimensional measure of the intended construct. The CR helps describe the percentage of data that matches the expected response pattern with values above 0.9 generally recommended for interpreting data as unidimensional. In calculating scale errors for our data, we use the Goodenough–Edwards method (Goodenough, 1955). The value from MMR helps in interpreting the CR by considering how many expected response patterns would occur by chance with lower values being desired for interpretation. The PI is simply the difference between CR and MMR with larger values indicating more evidence of validity to interpretation of scores. Finally, the coefficient of scalability is similar to CR, but the observed errors are divided by the marginal errors instead of total responses. Values above 0.6 are generally recommended for CS.

In order to test if our instrument was capable of seeing differences among our populations, we sought to compare universities based on factor scores. To make this comparison, we used an ANOVA to see if a measure showed significant difference among our sampled universities. In cases where we observed a significant difference, we followed ANOVA with a Tukey test to determine which universities were statistically different. Additionally, Cohen's f was calculated as a measure of effect size to understand the meaningfulness of these differences (Chen & Chen, 2012). All ANOVA procedures were performed in R (R Core Team, 2021).

In addition to seeing if our data showed differences among universities, we wanted to understand how these

scales we measured related to the outcome of EBIP adoption. In order to accomplish this goal, we performed a regression analysis in R with each measured construct predicting EBIP adoption. Additionally, to test hypotheses relating to the impact of social networks in teaching practice, we also used indegree, a measure of opinion leadership (Valente & Pumpuang, 2007), to test whether faculty who are seen as opinion leaders are themselves using EBIPs. The regression with and without indegree were run separately to evaluate the utility of adding indegree into the regression model. Among the faculty who completed the instrument, indegree ranged from a minimum of zero to a maximum of 24 with a median value of 6.

Results

Testing assumptions

Before running factor analysis, the data were summarized by item descriptives (Additional file 2: Table S1), inter-item correlation tables (Additional file 2: Table S2), and item response rates (Additional file 2: Table S3) and tested for assumptions utilized in factor analysis (e.g., normality). Results from these summaries are included in the Additional file 2. The results warranted no changes to the data or analysis plan.

Exploratory factor analysis

An EFA of one-half of the data was used to determine the underlying factor structure of items in the instrument. For determining the appropriate number of factors, best practices suggest combining multiple methods in addition to comparison with originally established theory. For our analysis, we combined Kaiser's criterion, Scree analysis, and parallel analysis. Kaiser's criterion, parallel analysis, and our intended design converge on a three-factor instrument underlying the data while Scree analysis suggests a four-factor instrument is reasonable (Fig. 2). Using these findings, we calculated three- and four-factor EFA solutions. Upon inspecting the four-factor solution, it split the six items for strategic complements into two factors based on standard and reverse wording of items. We chose to keep three factors, in alignment with our intended design, Kaiser's criterion, and parallel analysis.

Item pattern loadings into the three-factor solution are summarized in Additional file 2: Table S4, as are the factor correlations (Additional file 2: Table S5) for this model. These initial results are generally consistent with our intended design; however, one item intended to describe climate did not load into any factor. This item was worded in a way unlike the other climate items starting with "my teaching is…" while the rest of the items to address climate started with "the campus culture is…". Therefore, we decided it was best
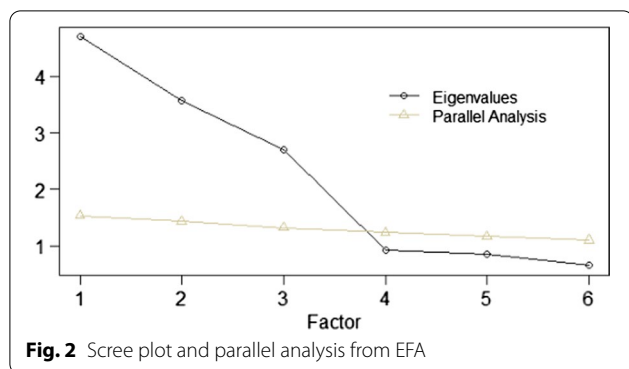
Fig. 2 Scree plot and parallel analysis from EFA

to remove the item and rerun the EFA. After repeating the analysis with this item removed, we observed a similar three-factor solution. Item pattern loadings (Table 1) are generally strong within a single factor with no substantive cross-loading. Additionally, the factors show insignificant or small correlations with each other (Table 2). Based on our intended design, we named the factors "perception of strategic complements", "perception of interdependence", and "perception of campus climate toward teaching". For simplicity, the factors will be referred to as "complements", "interdependence", and "climate", respectively.

Confirmatory factor analysis

The factor structure suggested by EFA was then used to conduct a CFA on the reserved portion of the data.

Table 1 Pattern loadings for EFA, item loadings < 0.4 suppressed, $N=150$

Item	Complements	Interdependence	Climate
comp1	0.876		
comp2	0.805		
comp3	0.603		
comp4	0.757		
comp5	0.545		
comp6	0.718		
interd1		0.762	
interd2		0.767	
interd3		0.589	
interd4		0.580	
interd5		0.853	
interd6		0.858	
clim2			0.782
clim3			0.855
clim4			0.699
clim5			0.868
clim6			0.570

Table 2 Correlation of factors from the EFA

Item	1	2	3
1. Complements			
2. Interdependence	0.138		
3. Climate	0.146	0.284*	

*$p < 0.05$

CFA allows us to determine if the reserved data are consistent with a proposed model by analyzing the data with a predefined model. Without any modifications to the model, the fit statistics from this CFA were χ^2 (116, $N=150$) $=303.297$, $p<0.0001$; RMSEA $=0.104$; CFI $=0.822$; and SRMR $=0.072$ (see Additional file 2: Fig. S1 for the structure and standardized loadings for this model). The χ^2 indicates significant misfit between the proposed model and observed data, and the other fit statistics are not within generally accepted levels (RMSEA <0.06, CFI >0.95, and SRMR <0.08; (Hu & Bentler, 1999)) indicating the proposed model is not an adequate approximation of the observed data.

One method to improve model fit is using a bifactor model as it can model instrument level effects in the measurement (Brown, 2015; Xu et al., 2016). When subjecting the current data to a bifactor analysis, the calculation failed to converge and increasing iterations beyond default levels simply resulted in greater divergence. This behavior was observed in both Mplus and R suggesting that the bifactor model is a poor fit for our data.

To account for misfit in the model, the potential for correlated error among the items was evaluated. These correlated error terms allow us to explain variance among items that goes beyond what is suggested by the factor structure. Based on modification indices from Mplus and inspection of individual items, two sources of correlated error within the Interdependence factor were proposed. First, the items were written in pairs which contained similar wording but with opposing sets of adjectives (good/bad, succeed/fail, and gain/loss). Similarly worded items are a common source of correlated error (Brown, 2015), and these three correlated errors were included in the model. Additionally, within the Interdependence factor, three items have standard wording (good, succeed, gain) while the other three have reverse wording (bad, fail, loss). Sets of items with reverse wordings have been known to have correlated error with each other (Brown, 2015). Therefore, correlated error was also modeled among the reverse-worded items (bad, fail, and loss) accounting for three more error terms. Modification indices and theoretical considerations for other pairs of items were not large enough to support further restricting the model.

The standardized model produced after these modifications is shown in Fig. 3. All paths show strong standardized loading coefficients (0.427–0.878). The model produced a significant χ^2 (110, $n=150)=152.580$, $p=0.0045$ indicating significant misfit, but less misfit than previously seen. Comparing to the earlier model which produced χ^2 (116, $N=150)=303.297$, we see that the change in χ^2, 150.717, over 6 degrees of freedom demonstrates a large, statistically significant fit improvement ($p<0.0001$). Using the other fit indices, we observe RMSEA $=0.051$, CFI $=0.960$, and SRMR $=0.067$ which indicate that our model produces a reasonable approximation of the observed data based on accepted standards (Hu & Bentler, 1999). Additionally, the factors are not strongly correlated to each other (see labeled double-headed arrows between latent constructs in Fig. 3) suggesting the scales are measuring independent constructs.

Internal consistency

Internal consistency is a description of how well different items intended to describe the same measure relate to each other. Table 3 reports values for Cronbach's alpha, Omega, and Raykov's rho to show evidence for internal consistency. Cronbach's alpha suggests good internal consistency (0.785–0.802), and Omega (0.694–0.816) and Raykov's rho (0.696–0.816) similarly support that the scales are reasonable with improvement on the complements and climate scales and a decrease in the interdependence scale.

The fit of the model suggests that it is reasonable to produce factor scores for interpretation. There are two standard methods of computing scores from items in a factor. A simple version is to simply take an average of each item in the factor. A more sophisticated method is to use the measurement model to estimate values based on the loading of each item and correlation among terms in the model. In addition to the ease of computation, the advantage to a simple average is that it can easily be compared to the scale that produced it to help interpret the score. So on a 7-point scale, an average of 4 would represent a neutral attitude with <4 being below neutral and >4 being greater than neutral. These averages also allow for future researchers to make direct comparisons to previously published data.

In contrast, using factor scores produced by the measurement model allows for more meaningful relative comparisons among individuals. This advantage for comparison comes in part from the fact that factor scores are calculated based on the difference from the average response relative to the standard deviation of responses. Factor scores are also helpful in that they can model effects of items contributing to the factor in different weights and compute the removal of measurement error terms. As both types of score (simple average and factor score) have value in different ways of looking at the data, both will be used in discussing the responses we received with care to mention which score is being used in a particular analysis. In the analysis and visualization of typical responses, the scores are based on simple average responses. These values are used as their simplicity in calculation makes the results more accessible to a broader audience of potential users of the CAFI. Additionally, the values are easier to interpret in terms of the construct as the alternative of basing the score on the average response would limit interpretation into how faculty members compare to each other and not how the construct is perceived on the campus as a whole. In contrast, ANOVA and regression analysis will use computed factor scores as in these cases, it is the differences among individual faculty members that we seek to explore.

EBIP adoption

To support the reliability of the Guttman scale to measure EBIP adoption, we explored a range of statistics designed to support the ability to interpret the data as unidimensional (McIver & Carmines, 2011). These statistics; CR, MMR, PI, and CS; along with their method of calculation are presented in Table 4. From these data, we have evidence that our items are functioning as a unidimensional scale, and it is appropriate to interpret scores as such.

Analyzing scores from the CAFI

Figure 4 shows the box plot of simple average responses to all the factors and the EBIP adoption score. This visualization is chosen to help visualize the range of responses and potentially see any outliers. For EBIP adoption, we see that the median is at or above 4 on each campus with Uni1 appearing to have higher adoption on average.

In the Complements factor, the average across all participants is 5.10 which is close to a "Somewhat agree" response. In terms of the range of responses, we see they tend to vary from 3 (somewhat disagree) to 7 (strongly agree). There appears to be variation across the three sampled universities while the median faculty is in all cases above a neutral response of 4. For the Interdependence factor, the average score of 5.77 indicates a fairly positive response suggesting feelings of cooperation and trust among faculty. In terms of the range of responses, scores tend to range from 4 (neutral) all the way to 7 (strong interdependence). Among the sampled universities, there does not appear to be a strong difference is the distribution of interdependence scores. In the Climate factor, the average score is a 4.74. On this scale, we see the full range of responses with faculty from Uni3 reporting the lowest possible score of 1 while faculty from all
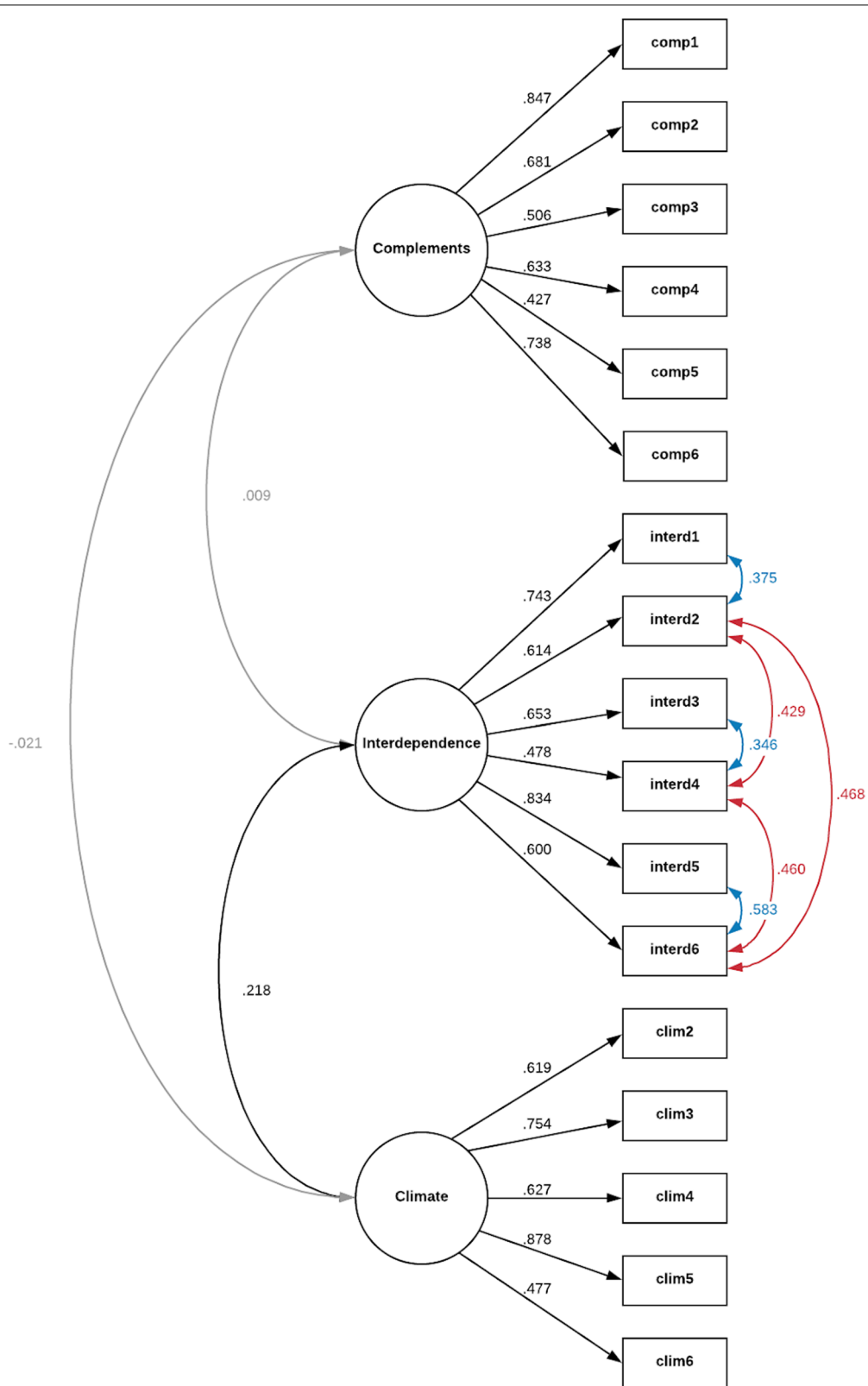
Fig. 3 Standardized CFA. Gray arrows represent non-significant paths (> 0.05), blue arrows represent correlated error between similarly worded pairs, red arrows represent correlated error among reverse-worded items

Table 3 Measures of internal consistency

Factor	Cronbach's alpha	Omega	Raykov's rho
Complements	0.785	0.805	0.794
Interdependence	0.873	0.694	0.696
Climate	0.802	0.816	0.816

Table 4 Summary of statistics to support reliability of EBIP adoption scale

Statistic	Formula	Observed	Recommended thresholds
CR	(TR − SE)/TR	0.975	> 0.9
MMR	(TR − ME)/TR	0.703	Lower better
PI	CR − MMR	0.273	Larger better
CS	(ME − SE)/ME	0.917	> 0.6

CR coefficient of reproducibility, *MMR* minimal marginal reproducibility, *PI* percent improvement, *CS* coefficient of scalability, *TR* total responses, *SE* scale errors, *ME* marginal errors

universities reported the highest possible score of 7. Comparing the universities in our sample, we see that Uni1 has a median above a 5 while the others are lower than 5 but above 4 suggesting some difference in perceptions of climate. These climate scores also have a wide range both below and above the neutral response of 4.

One observation from these boxplots is that faculty at Uni2 had a few outlier responses and situations where the median and 75 percentile responses were identical (represented visually with no 'box' to the right of the median line). Based on this observation, regression analyses were run both with and without Uni2 due to concerns in how the distribution of responses might affect analysis. The results presented will highlight those that came from analyzing all three campuses as a whole while commenting whenever the removal of Uni2 resulted in a change in interpretation of results.

Institutional comparisons

ANOVA was then used to determine if any of the possible differences in responses reached levels of statistical significance. For each ANOVA that demonstrated evidence of significance difference, subsequent Tukey testing to determine the source for the difference is summarized in Table 5. For EBIP adoption, we see a significant difference among universities. For effect size, this difference has a Cohen's f of 0.22, close to a medium effect size, which suggests that this difference can be fairly substantial (Chen & Chen, 2012). After the ANOVA, follow-up Tukey tests show Uni1 has higher adoption than the others while there was not a significant difference between Uni2 and Uni3.

For the factor scores, variation was seen on the observation of significant difference and the magnitude of that difference. Here, we see that the Climate scale shows one of the largest differences among the universities. ANOVA showed evidence of significant difference with an effect size (Cohen's f) of 0.26. This medium effect size suggests this difference among these campuses is substantial. Follow-up Tukey tests show that Uni3 had a lower Climate score than the other campuses with no evidence of significant difference between Uni1 and Uni2.
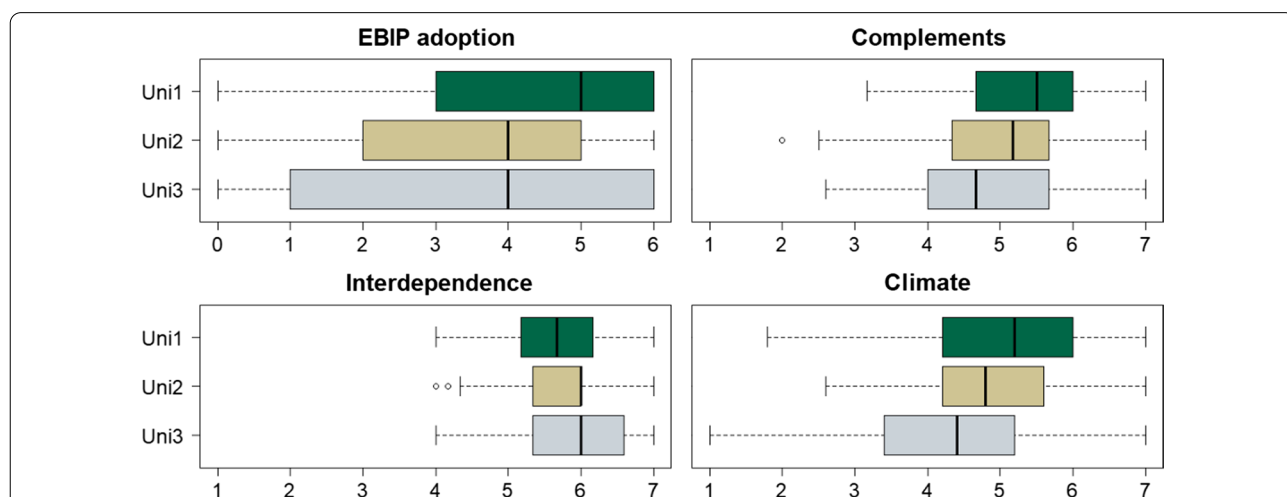
Fig. 4 Box plots of average item responses to each scale. Outliers (individuals with scores separated from either end of the interquartile range by a distance more than 1.5 times the length of interquartile range) are represented by dots

Table 5 Tukey results indicating significance of pairwise comparisons

Variable	Measured	Pairwise comparison		Mean diff	Std error	*p* value
University	EBIP adoption	Uni1	Uni2	0.70	0.34	0.037
		Uni1	Uni3	1.04	0.34	0.001
		Uni2	Uni3	0.34	0.34	0.457
	Complements	Uni1	Uni2	0.36	0.17	0.023
		Uni1	Uni3	0.37	0.16	0.017
		Uni2	Uni3	0.01	0.16	0.997
	Climate	Uni1	Uni2	0.19	0.19	0.470
		Uni1	Uni3	0.66	0.19	< 0.001
		Uni2	Uni3	0.48	0.19	0.007

ANOVA for interdependence factor did not provide evidence of significant difference

The next largest difference observed would be for the complements factor. This difference was significant and had a Cohen's f of 0.18 (between small and medium effect size). Follow-up Tukey tests show Uni1 has higher complements than Uni2 or Uni3 while there is not a significant difference between Uni2 and Uni3.

For interdependence, we do not see a significant difference among universities from an ANOVA. The difference that is observed has a Cohen's f of 0.08 (small) which supports the similarity among the three institutions on this scale.

Regression analysis

Another consideration for our theory-based instrument is how the three scales relate to our outcome of interest, namely EBIP adoption. Regression analysis allows us to examine how each scale predicts EBIP adoption compared to the other scales. Furthermore, we want to examine the robustness of scale prediction when opinion leadership is also a predictor of EBIP adoption in the model. As noted earlier, opinion leadership is measured by the number of times a faculty member is nominated by other faculty as a teaching discussion partner (referred to as indegree), with higher values suggesting the individual is an opinion leader whose ideas are valued. Hence, we performed a regression analysis with the computed factor scores of complements, interdependence, and climate and with and without indegree predicting the score on the EBIP adoption scale.

The results of the two regressions are in Table 6. By comparing the nested models, with and without indegree as a predictor, we see that adding indegree produces a significant improvement to the model. The important point, however, is that the CAFI scale effects are robust even when indegree is introduced as a predictor. The significance of indegree produces an additional finding, worth future research attention, that faculty who receive more teaching discussion ties are more likely to be EBIP users than faculty who receive fewer.

Looking at the other terms from the regression, we see a significant effect that higher scores on the complements scale predict higher levels of adoption of EBIPs. We also find climate to be a significant, but negative, predictor of EBIP adoption. However, this finding of significance is not robust to removal of Uni2 from the regression as running the model with just Uni1 and Uni3 results in a significance of 0.078 which is outside the conventional range of < 0.05.

Table 6 Comparison of regression analysis predicting EBIP adoption score based on factor scores both with and without indegree

Model		*b*	Std error	Beta	*t*	Sig	Fit	Difference
1	(Intercept)	3.87	0.10		40.353	< 0.001**		
	Complements	1.20	0.10	0.58	12.044	< 0.001**		
	Interdependence	− 0.10	0.15	− 0.03	-0.681	0.496		
	Climate	− 0.16	0.09	− 0.09	-1.832	0.068	$R^2 = 0.333$**	
2	(Intercept)	3.19	0.19		17.108	< 0.001**		
	Complements	1.06	0.10	0.51	10.414	< 0.001**		
	Interdependence	− 0.13	0.15	− 0.04	-0.888	0.375		
	Climate	− 0.17	0.09	− 0.10	-1.973	0.049*		
	Indegree	0.10	0.02	0.21	4.258	< 0.001**	$R^2 = 0.372$**	$\Delta R^2 = 0.039$**

b represents unstandardized regression weight, and beta indicates standardized regression weights

*indicates $p < 0.05$. **indicates $p < 0.01$

Discussion

Validity and internal structure

We developed the CAFI to measure factors related to institutional change in STEM education based on our understanding of the CACAO model. We consider faculty's perception of the relative advantage of taking the same action as their colleagues (strategic complements), feelings of trust and altruism within a department (interdependence), and perceptions of the institutional attitudes toward teaching (climate). In order to support any conclusions based on analysis of scores produced by the CAFI, we sought to collect evidence of validity and evidence that the scales can show variance in a population (AERA 2014; Arjoon et al., 2013). One source of validity evidence is content validity which looks for consistency between the intended construct and the items used to measure it. For the CAFI, scale items started with a foundation in academic literature as described in the introduction. Additionally, our use of an external expert panel further supports the content validity of our instrument by providing an outsider perspective (AERA 2014).

Another aspect of evidence for an instrument is internal structure, which relates to how well similarly themed items covary. To support the internal structure of this instrument, we performed EFA and CFA on separate data sets. The EFA along with our established change theory provided strong evidence of a three-factor structure with all but one item loading onto its respective factor. The single item was removed from the analysis and repeating the EFA without the single item resulted in strong evidence for the structure that divided items into factors related to strategic complements, interdependence, and climate. This result suggests that the three-factor solution is an appropriate and parsimonious way to describe our data.

This EFA was followed by CFA on a reserved portion of the data to see how well the reserved data matched the predefined model. The initial CFA showed promise but did not reach the accepted standards desired for factor structure fit in measures (Hu & Bentler, 1999). By modeling correlated errors between some similarly worded items and among reverse-worded items on the interdependence scale, we saw improvements in the fit indices reaching to values that meet the typically accepted thresholds suggesting the model is a reasonable approximation of our data. These findings support the internal structure of the CAFI and give us evidence that it is appropriate to calculate factor scores (Brown, 2015).

Examination of scores from scales

Within the CACAO model, adoption of an innovative practice is separated into a series of discrete steps (Dormant & Lee, 2011). From the scale for EBIP adoption implemented in this study, a score of zero represents preawareness, 1–3 represent levels of knowledge, and scores 4–6 represent levels of adoption (Landrum et al., 2017; Viskupic et al., 2022). Looking at the median scores in each university, we observed a value greater than or equal to 4 suggesting that most survey respondents used EBIPs. This is a promising finding as all three universities in this sample engaged in campus-wide activities to promote adoption of EBIPs among STEM faculty. It is also noteworthy to consider that a tenth of the respondents (~ 30) had a score of zero on this scale suggesting these faculty are not aware of EBIPs. Based on this finding, the instrument can be used to directly detect if institutional change initiatives to spread awareness and use of EBIPs are having the intended effect.

In terms of using these scores for EBIP adoption to compare universities, ANOVA showed that among our 3 sampled universities, we saw a statistically significant difference in EBIP adoption. This finding helps support that the instrument can detect differences, and the scale does not produce the same scores across different locations. One concern with self-report data is that faculty might feel pressure to report higher scores than are reflected in actual practice. Previous research has supported that instructor self-report, student feedback, and observation tend to produce similar results (Durham et al., 2018), suggesting that any one of these types of data can reflect actual practice. Additionally, none of the results from this survey were shared back to university or department in an identifiable way, lessening any concern for faculty to not respond honestly. Therefore, we believe the scores we report are representative of our surveyed departments, but care should be used if this scale is administered in a setting that might encourage inflated responses.

On the strategic complements scale, the average response of 5.10 on the 7-point scale suggests that faculty in our sample tend to view the adoption of EBIPs as a mutually reinforcing behavior (Jackson & Zenou, 2015). This finding would support the idea that as more faculty adopt EBIPs, other faculty will follow among those surveyed in our sampled population. In terms of the CACAO model, this finding would help provide evidence that as more faculty members adopt EBIPs, the relative advantage of an individual faculty member also adopting will go up and lead to more adoption. As the perception that EBIPs are valuable increases, this would increase the likelihood that they are adopted. This finding suggests that over time, EBIPs can spread in these departments as the increase in relative advantage leads to more adoption which leads to even more strategic complementarity. Based on ANOVA, we see a statistically significant difference in complements among our sampled departments,

suggesting that the CAFI will be an effective tool in comparisons. Another finding is that we do observe that the university which reported a statistically significant higher complements score is also the university with statistically significant higher EBIP adoption.

For the interdependence scale, the average response of 5.77 suggests that faculty generally have a view that their fates are intertwined (Aktipis et al., 2018). Based on this high value from these sampled universities, we have evidence that there is a degree of trust among faculty (Boon & Holmes, 1991) and that discussing or using EBIPs is unlikely to have a negative social impact (Dormant & Lee, 2011). With a finding like this, a change agent implementing a change initiative can use this interdependence as leverage when designing activities and policies. For example, change agents can consider activities such as faculty retreats or team teaching that require mutual effort. Another option for change agents is to highlight the success of faculty through team awards and recognitions rather than individual awards. Based on ANOVA, we fail to see evidence of a statistically significant difference when comparing these universities. Since all responses were generally high in our sample, it would be interesting to see data on this scale from settings where interdependence might be expected to be lower. For example, a university where departments are undergoing structural changes, such as being newly created from formerly separate units. Another setting where perception of interdependence may be lower is within institutions that implement competitive 'merit pay' structures, as success of a colleague could be seen as having a negative impact on the availability of merit pay for other faculty members.

The climate scale responses suggest that faculty have a wide range of perceptions of their institution's readiness for change. While the average response of 4.74 on the 7-point scale is above a neutral response, several faculty members reported the theoretical minimum of 1 while others viewed climate very positively, reaching the theoretical maximum of 7. This finding suggests that there is a wide range of the perception of the compatibility between implementing EBIPs and the support from the university in implementing EBIPs. As compatibility can influence potential adopters to use an innovative practice like EBIPs (Dormant & Lee, 2011), this finding would suggest that change agents should consider efforts to investigate why there are different perceptions of climate and encourage those with positive perceptions to share their experiences. Leadership can be important in communicating and leading this effort to support the use of innovative teaching practices (Walter et al., 2021). Based on ANOVA, the CAFI can detect a statistically significant difference in climate among our sampled universities,

indicating that our instrument can be useful in making comparisons.

Relating scores to EBIP adoption

In the regression analysis predicting EBIP adoption from complements, interdependence, climate, and indegree, we obtained some initial insights into the predictive power of the CAFI in describing EBIP adoption. When comparing regressions that include and exclude indegree, we see that the inclusion of indegree results in a statistically significant improvement of fit in the regression. As indegree serves as a measure for opinion leadership (Valente & Pumpuang, 2007), this finding supports that the opinion leaders in the teaching discussion networks are themselves users of EBIPs. Therefore, training some faculty members to spread EBIPs in a department can be an effective strategy to spread the adoption of EBIPs. Andrews et al. (2016) found that DBER faculty were quite influential in their networks, which would support the spread of EBIPs. Our team has previously reported an exponential random graph model analysis providing evidence that EBIP users tend to seek out other EBIP users for teaching discussions (Lane et al., 2020). The result from our regression here is consistent with this previous finding in that faculty who use EBIPs are more likely to be sought as discussion partners for conversations about teaching. This result is also consistent with previous findings that faculty members with more extensive networks also showed more learner-centered teaching behavior (Middleton et al., 2022).

Once we established that indegree was predictive in our model, we focused on the relationship between the other scales on the CAFI and EBIP adoption. Among the findings is evidence of a direct relationship between the complements and climate with EBIP adoption. As the regression coefficient for complements is positive and significant, we have support that the perception that use of EBIPs provides a shared benefit is predictive of faculty members using EBIPs. This adds to our previous finding that the perception of strategic complementarity was generally high at our sampled universities by saying the perception is also connected to the use of EBIPs. Therefore, we have evidence that promoting the use of EBIPs as mutually beneficial may support their adoption among other faculty members.

For the regression coefficient between climate and EBIP adoption, we find the value is both negative and significant when looking at the results from the model which took all campuses into account. Based on this result, we would expect that an increased perception that the university supports change in teaching practice results in less use of EBIPs. However, as our research design is correlational rather than causal, it could also be that a decision

not to adopt EBIPs is what leads to the higher perception of the campus climate toward EBIP use. To better understand this negative correlation between EBIP adoption and climate, we ran a correlation test of EBIP adoption score to climate when splitting the data between faculty who do not use EBIPs (adoption score < 4) and those who do use EBIPs (adoption score ≥ 4). We see the correlation change from − 0.236 for nonusers to 0.055 for users. A potential reason that we observe this tendency could be that faculty members who are not using EBIPs perceive the campus as supporting their use but decide against adopting EBIPs for other reasons.

As of now, there is no significant relationship between EBIP adoption and the interdependence scale. This suggests that feelings of trust among faculty are not directly predicting EBIP adoption among our sampled universities. It is possible that even if interdependence does not have a direct impact on EBIP adoption, it can have an indirect effect through how people form social connections and how they persist. Additionally, as our sample showed little variance on the interdependence scale, it is possible that any effect it has may just not be apparent, and collection of data in a setting with a broader range of responses may provide more insight into the impact of interdependence on teaching practice.

Implications

We have shown how using the CAFI can quickly characterize departments or universities. This instrument provides change agents with a tool to characterize local conditions on their respective campuses and other campuses that have used it. The survey is short enough to be completed in less than 10 min, and we observed reasonable response rates even with little or no formal incentive. Based on our results, we see that campuses and individuals have a range of responses on many of these scales, and the scales have a direct relationship to EBIP adoption.

For those researching the impact and process of change initiatives, this instrument provides measures for analyzing a population. By considering the progression of EBIP adoption and the factor scores from this instrument based on the components of an ideal change as presented in the CACAO model, a researcher can develop descriptive and predictive summaries of different initiatives. The instrument produces scores with decent variance and good internal structure characteristics. The instrument can be provided at the beginning of a project, at intermediate steps, and afterwards to have a longitudinal comparison. The present study focused on institutions in later stages of reform initiatives, so it would also be of value to see how this instrument functions in settings that are earlier in the process of institutional change.

While we have not yet demonstrated how these scales behave within the context of a developing change project, we have developed a tool that can be used to empirically monitor constructs that theory tells us might influence a change process. We hope that groups planning to engage in a change effort in higher education will consider administering this instrument at a very early stage, something which we were unable to do, and then at regular intervals aligned with the implementation of their specific set of change strategies. The initial empirical data will provide a mechanism to develop in situ hypotheses about change readiness and change strategies that can be tested via subsequent data collections as a change effort continues. The first change leadership teams who engage with the instrument in this way will also provide invaluable information about its successes and failures, which should serve to highlight both the strengths of the current form of the instrument and to indicate potential improvements, which are an important part of any instrument's ultimate utility.

For the practical work of individual change agents themselves, our results documenting a link between EBIP use and the complements scale, in particular, demonstrate that individual change agents are likely to benefit from knowing the degree to which a specific department (or other unit that is to be a primary site of change) perceives mutual benefits of EBIPs. Activities designed to convince faculty of the mutual benefit of implementing a pedagogical approach are quite different from, for example, activities designed to assist faculty with learning how to implement that pedagogical approach effectively. Change agents do not have infinite time, and individual change agents have different strengths. Having greater insight about where to place emphasis can help change agents make the best use of their time and skills.

Limitations

There are some concerns that should be considered while interpreting these results. While we did target five disciplines at three different universities, the results may or may not be transferable to other institutions or academic departments. This concern is particularly true for universities that do not have large enrollments or strong research emphasis. A similar concern arises from the fact that the departments in this study had different sizes and response rates. It is possible that some of the observed differences in response rate could be attributed to available incentives. Aligning incentive structures across the institutions might have produced a uniform response rate. Regardless, more iterations at different types of institutions and academic departments would be helpful to support the evidence of validity. Additionally, while

exploratory and confirmatory factor analyses were performed on responses from different individuals, the responses were from the same departments and the same universities. This decision could result in a factor structure too specific to this particular case. As the instrument begins to be used more broadly, the factor structure should be examined before interpretation. For the CFA presented in this paper, addition of correlated error terms to the model allowed the model to reach accepted levels of fit but were proposed post hoc. There are also limitations inherent to our choice to develop a quantitative tool. This approach to data collection, while allowing someone to characterize a large population in a short time, can lack the depth and nuance that would be provided from other data collection methods such as semi-structured interviews. While we hope that others find the CAFI useful, we believe it should not preclude the collection of multiple forms of data in the effort to understand complex institutional changes. Finally, when examining the results from regression analysis, the research design does not allow for causal claims, so findings should be interpreted cautiously. The relationships are correlational, and further investigation is required to demonstrate a true causal link.

Conclusion

This report details efforts to develop and administer a new instrument, the CAFI, to better understand institutional change within university STEM departments. This instrument uses the CACAO model of change to link the intended outcome of EBIP adoption with perception of EBIPs as mutually reinforcing (strategic complements), perception of faculty having their fates intertwined (interdependence), and perception of institutional readiness for change (climate). Through discussions with an external advisory board for content validity and factor analysis for internal structure, we have support for the validity of interpreting scores produced by the CAFI. We demonstrated through ANOVA that many of these scales show significant differences across different universities, supporting the utility of the instrument in making comparisons. We also found by regression analysis that the complements and climate scale along with status as an opinion leader in a department have a significant relationship to the outcome of EBIP adoption. The CAFI serves as a tool to help change agents understand their campus environment and help support the implementation of EBIPs.

Abbreviations

ANOVA: Analysis of variance; CACAO: Change–Adopter–Change Agents–Organization; CAFI: Cooperative Adoption Factors Instrument; CFA: Confirmatory factor analysis; CR: Coefficient of reproducibility; CS: Coefficient of scalability; EBIP: Evidence-based instructional practices; EFA: Exploratory factor analysis; MMR: Minimal marginal reproducibility; PI: Percent improvement.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40594-022-00364-w.

> **Additional file 1.** The text of the Cooperative Adoption Factors Instrument as seen by respondents.
>
> **Additional file 2.** Information regarding simple descriptive statistics, missing data, item correlations, initial exploratory factor analysis results, and initial confirmatory factor analysis results. Supplementary tables and figures.

Declarations

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

Author details
[1]Chemistry, University of South Florida, 4202 E Fowler Avenue CHE201, Tampa, FL 33620, USA. [2]Center for the Improvement of Teaching and Research in Undergraduate STEM Education, University of South Florida, Tampa, USA. [3]Anthropology, Boise State University, Boise, USA. [4]Sociology, University of South Florida, Tampa, USA. [5]School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, USA. [6]Office of the Provost, Boise State University, Boise, USA. [7]Department of Biology Teaching and Learning, University of Minnesota, Minneapolis, USA. [8]Ecology, Evolution, and Behavior, Boise State University, Boise, USA. [9]Integrative Biology, University of South Florida, Tampa, USA. [10]Chemistry and Biochemistry, Boise State University, Boise, USA. [11]Chemistry, University of Virginia, Charlottesville, USA.

References
Aktipis, A., Cronk, L., Alcock, J., Ayers, J. D., Baciu, C., Balliet, D., Boddy, A. M., Curry, O. S., Krems, J. A., Muñoz, A., Sullivan, D., Sznycer, D., Wilkinson, G. S., & Winfrey, P. (2018). Understanding cooperation through fitness interdependence. *Nature Human Behaviour, 2*(7), 429–431.
American Education Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, &

NCME). (2014). *Standards for educational and psychological testing. American Education Research Association* (4th ed.). AERA Publication Sales.

Anderson, S. E. (2014). Understanding teacher change: Revisiting the concerns based adoption model. *Curriculum Inquiry, 27*(3), 331–367.

Andrews, T. C., Conaway, E. P., Zhao, J., & Dolan, E. L. (2016). Colleagues as change agents: How department networks and opinion leaders influence teaching at a single research university. *CBE Life Sciences Education, 15*(2), ar15.

Arjoon, J. A., Xu, X., & Lewis, J. E. (2013). Understanding the state of the art for measurement in chemistry education research: Examining the psychometric evidence. *Journal of Chemical Education, 90*(5), 536–545.

Bollen, K. A. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review, 45*(3), 370–390.

Boon, S., & Holmes, J. (1991). The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. In R. A. Hinde & J. Groebel (Eds.), *Cooperation and prosocial behavior* (pp. 190–211). Cambridge University Press.

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

Brown, T. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

Bulow, J. I., Geanakoplos, J. D., & Klemperer, P. D. (1985). Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political Economy, 93*(3), 488–511.

Chen, S., & Chen, H. (2012). Cohen's f statistic. In N. J. Salkind (Ed.), *Encyclopedia of research design.* SAGE Publications Inc.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Dormant, D., & Lee, J. (2011). *The chocolate model of change.* Lulu Publishing.

Durham, M. F., Knight, J. K., Bremers, E. K., DeFreece, J. D., Paine, A. R., & Couch, B. A. (2018). Student, instructor, and observer agreement regarding frequencies of scientific teaching practices using the Measurement Instrument for Scientific Teaching-Observable (MISTO). *International Journal of STEM Education, 5*(1), 1–15.

Durham, M. F., Knight, J. K., & Couch, B. A. (2017). Measurement instrument for scientific teaching (MIST): A tool to measure the frequencies of research-based teaching practices in undergraduate science courses. *CBE Life Sciences Education, 16*(4), ar67.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415.

Gibbons, R. E., Laga, E. E., Leon, J., Villafañe, S. M., Stains, M., Murphy, K., & Raker, J. R. (2017). Chasm crossed? Clicker use in postsecondary chemistry education. *Journal of Chemical Education, 94*(5), 549–557.

Goodenough, W. H. (1955). A technique for scale analysis. *Educational and Psychological Measurement, 4*(3), 179–190.

Hall, G. E., & Hord, S. M. (1987). The concerns-based approach: An overview. *Change in schools: Facilitating the process* (pp. 1–22). Suny Press.

Hames, R., & McCabe, C. (2007). Meal sharing among the Ye'kwana. *Human Nature, 18*(1), 1–21.

Hardin, G. (1968). The tragedy of the commons. *Science, 162*(3859), 1243–1248.

Hayward, C. N., & Laursen, S. L. (2018). Supporting instructional change in mathematics: Using social network analysis to understand online support processes following professional development workshops. *International Journal of STEM Education, 5*(1), 1–19.

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

Hord, S. M., Rutherford, W. L., Huling-Austin, L., & Hall, G. E. (1987). *Taking charge of change.* Southwest Educational Development Laboratory.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Jackson, M. O., & Zenou, Y. (2015). Games on networks. *Handbook of Game Theory with Economic Applications, 4*(1), 95–163.

Kezar, A. (2014). Higher education change and social networks: A review of research. *The Journal of Higher Education, 85*(1), 91–125.

Knaub, A. V., Henderson, C., & Fisher, K. Q. (2018). Finding the leaders: An examination of social network analysis and leadership identification in STEM education change. *International Journal of STEM Education, 5*(1), 1–14.

Komperda, R., Pentecost, T. C., & Barbera, J. (2018). Moving beyond alpha: A primer on alternative sources of single-administration reliability evidence for quantitative chemistry education research. *Journal of Chemical Education, 95*(9), 1477–1491.

Kotter, J. P., & Cohen, D. S. (2002). The heart of change. *The heart of change: Real-life stories of how people change their organizations* (pp. 1–14). Harvard Business Review Press.

Landrum, R. E., Viskupic, K., Shadle, S. E., & Bullock, D. (2017). Assessing the STEM landscape: The current instructional climate survey and the evidence-based instructional practices adoption scale. *International Journal of STEM Education, 4*(1), 1–10.

Lane, A. K., McAlpin, J. D., Earl, B., Feola, S., Lewis, J. E., Mertens, K., Shadle, S. E., Skvoretz, J., Ziker, J. P., Couch, B. A., Prevost, L. B., & Stains, M. (2020). Innovative teaching knowledge stays with users. *Proceedings of the National Academy of Sciences, 117*(37), 22665–22667.

Lane, A. K., Skvoretz, J., Ziker, J. P., Couch, B. A., Earl, B., Lewis, J. E., McAlpin, J. D., Prevost, L. B., Shadle, S. E., & Stains, M. (2019). Investigating how faculty social networks and peer influence relate to knowledge and use of evidence-based teaching practices. *International Journal of STEM Education, 6*(1), 1–14.

Lewin, K. (1947). Frontiers in group dynamics: Concept, method and reality in social science; Social equilibria and social change. *Human Relations, 1*(1), 5–41.

López, N., Morgan, D. L., Hutchings, Q. R., & Davis, K. (2022). Revisiting critical STEM interventions: A literature review of STEM organizational learning. *International Journal of STEM Education, 9*(1), 1–14.

Ma, S., Herman, G. L., Tomkin, J. H., Mestre, J. P., & West, M. (2018). Spreading teaching innovations in social networks: The bridging role of mentors. *Journal for STEM Education Research, 1*(1), 60–84.

McConnell, M., Montplaisir, L., & Offerdahl, E. (2019). Meeting the conditions for diffusion of teaching innovations in a university STEM department. *Journal for STEM Education Research, 3*(1), 43–68.

McDonald, R. P. (2013). Test homogeneity, reliability, and generalizability. *Test theory: A unified treatment* (pp. 76–119). Psychology Press.

McIver, J., & Carmines, E. (2011). An introduction to Guttman scaling. *Unidimensional Scaling* (pp. 41–61). SAGE Publications Inc.

Mestre, J. P., Herman, G. L., Tomkin, J. H., & West, M. (2019). Keep your friends close and your colleagues nearby: The hidden ties that improve STEM education. *Change the Magazine of Higher Learning, 51*(1), 42–49.

Middleton, J. A., Krause, S., Judson, E., Ross, L., Culbertson, R., Hjelmstad, K. D., Hjelmstad, K. L., & Chen, Y.-C. (2022). A social network analysis of engineering faculty connections: Their impact on faculty student-centered attitudes and practices. *Education Sciences, 12*(2), 108.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide* (7th ed.). Muthén & Muthén.

Naef, M., & Schupp, J. (2009). Measuring trust: Experiments and surveys in contrast and combination. SOEPpaper No. 167.

Quardokus, K., & Henderson, C. (2015). Promoting instructional change: Using social network analysis to understand the informal structure of academic departments. *Higher Education, 70*(3), 315–335.

R Core Team. (2021). *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology, 54*(2), 315–323.

Reinholz, D. L., & Andrews, T. C. (2020). Change theory and theory of change: What's the difference anyway? *International Journal of STEM Education, 7*(1), 1–12.

Rogers, E. (2003). *Diffusion of innovations*. Free Press.

Sachmpazidi, D., Olmstead, A., Thompson, A. N., Henderson, C., & Beach, A. (2021). Team-based instructional change in undergraduate STEM: Characterizing effective faculty collaboration. *International Journal of STEM Education, 8*(1), 1–23.

Schelling, T. C. (1966). *Arms and influence*. Yale University Press.

Schroeder, C. M., Scott, T. P., Toison, H., Huang, T. Y., & Lee, Y. H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching, 44*(10), 1436–1460.

Shadle, S. E., Liu, Y., Lewis, J. E., & Minderhout, V. (2018). Building a community of transformation and a social network analysis of The POGIL Project. *Innovative Higher Education, 43*(6), 475–490.

Shadle, S. E., Marker, A., & Earl, B. (2017). Faculty drivers and barriers: Laying the groundwork for undergraduate STEM education reform in academic departments. *International Journal of STEM Education, 4*(1), 1–13.

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Eagan, M. K., Esson, J. M., Knight, J. K., Laski, F. A., Levis-Fitzgerald, M., Lee, C. J., Lo, S. M., McDonnell, L. M., McKay, T. A., Michelotti, N., Musgrove, A., Palmer, M. S., Plank, K. M., … Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science, 359*(6383), 1468–1470.

Sturtevant, H., & Wheeler, L. (2019). The STEM Faculty Instructional Barriers and Identity Survey (FIBIS): Development and exploratory results. *International Journal of STEM Education, 6*(1), 1–22.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53.

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Nicole Arroyo, E., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., … Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences, 117*(12), 6476–6483.

Valente, T. W., & Pumpuang, P. (2007). Identifying opinion leaders to promote behavior change. *Health Education & Behavior, 34*(6), 881–896.

Viskupic, K., Earl, B., & Shadle, S. E. (2022). Adapting the CACAO model to support higher education STEM teaching reform. *International Journal of STEM Education, 9*(1), 1–20.

von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior (sixtieth-anniversary edition)*. Princeton University Press.

Walter, E. M., Beach, A. L., Henderson, C., Williams, C. T., & Ceballos-Madrigal, I. (2021). Understanding conditions for teaching innovation in postsecondary education: Development and validation of the Survey of Climate for Instructional Improvement (SCII). *International Journal of Technology in Education, 4*(2), 166–199.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*(6), 913–934.

Xu, X., Kim, E. S., & Lewis, J. E. (2016). Sex difference in spatial ability for college students and exploration of measurement invariance. *Learning and Individual Differences, 45*, 176–184.

Ziker, J., & Schnegg, M. (2005). Food sharing at meals. *Human Nature, 16*(2), 178–210.

Publisher's Note