RESEARCH   **Open Access**

Development of the Departmental Climate around Teaching (DCaT) survey: neither psychological collective climate nor departmental collective climate predicts STEM faculty's instructional practices

Lu Shi and Marilyne Stains*

Abstract

Background: Investigations into drivers and barriers to the uptake of learner-centered instructional practices in STEM courses have identified the climate within a department as a potential influential factor. However, few studies have explored the relationship between adoption of learner-centered instructional practices and departmental climate around teaching. Moreover, surveys that have been designed to measure climate around teaching have been focused on measuring individual faculty member's description of their colleagues' perceptions of the climate within their department (psychological collective climate) and ignored whether there was a consensus among respondents within the same department on these descriptions. This latter measure (departmental collective climate) is best aligned with the definition of organizational climate. There is thus a need to explore whether departmental climate measured at the individual or collective level relate to the use of learner-centered instructional practices.

Results: This study demonstrates that the Departmental Climate around Teaching (DCaT) survey provides valid and reliable data that can be used to measure psychological collective climate within a STEM department. Analysis of the 166 faculty members who responded to the survey indicated that (1) four different types of psychological collective climate existed among our population and (2) multiple types could be present within the same STEM department. Moreover, it showed that measuring departmental collective climate is challenging as few constructs measured by the DCaT survey reached high level of consensus within faculty members from the same department. Finally, the analysis found no relationship between psychological collective climate and the level of use of learner-centered instructional practices.

---

* Correspondence: mstains@virginia.edu
Physical Life Sciences Building, 409 McCormick Road, Charlottesville, VA
22904-4319, USA

Conclusions: Results from the validation studies conducted on the DCaT survey that most elements that define a climate (e.g., policies, practices, expectations) are lacking when it comes to teaching. These findings could explain the challenges experienced in this study in measuring departmental collective climate. Without these climate elements, faculty members are left to work autonomously with little expectations for growth in their instructional practices. Establishing policies, practices, and expectations with respect to teaching is thus an essential step toward instructional change at a departmental level.

Keywords: Departmental climate, Teaching climate, Higher education, STEM faculty, Evidence-based instructional practices, Instructional change

Introduction

A wave of instructional reforms within the last decade has focused on propagating learner-centered instructional practices in Science, Technology, Engineering, and Mathematics (STEM) courses at the postsecondary level. However, uptake has not reached the desired level (Stains et al., 2018), and efforts have been ongoing to identify levels (e.g., national, institutional, departmental, individual) and levers (e.g., promotion and tenure guidelines, evaluation of teaching, professional development opportunities) that would increase the pace of uptake. Multiple studies have demonstrated the complexity of STEM instructors' working environment (Anderson et al., 2011; Austin, 2011; Brownell & Tanner, 2012; Childs, 2009; Froyd, 2011; Gess-Newsome et al., 2003; Henderson & Dancy, 2007, 2011; Hora, 2012; Lund & Stains, 2015; Walczyk et al., 2007) and highlighted the importance of taking a system approach to promote instructional change (Austin, 2011; Corbo et al., 2016; Elrod & Kezar, 2016; The Coalition for Reform of Undergraduate STEM Education, 2014). Departments have been recognized as a key level of the system to target and several recent efforts and frameworks aim to explore approaches to promote change at this level (Austin, 2011; Corbo et al., 2016; Reinholz et al., 2017; Reinholz & Apkarian, 2018; The Coalition for Reform of Undergraduate STEM Education, 2014; Wieman et al., 2010). One characteristic of a department that has been advanced as critical to address is its climate around teaching (i.e., perceptions of policies, practices, and expected behaviors related to teaching).

Several studies have found that STEM faculty point to departmental climate around teaching as a barrier to instructional change (e.g., Henderson & Dancy, 2007; Shadle et al., 2017; Sturtevant & Wheeler, 2019). For example, these three studies found that faculty cited departmental norms defined by lecture-focused teaching as a barrier to using learner-centered instructional practices. Interestingly, Shadle et al. (2017) also surveyed STEM faculty at the authors' institution about drivers to instructional change. In alignment with other studies (Bouwma-Gearhart, 2012; Wieman et al., 2013), they found departmental-level drivers for instructional

change such as having discussions about teaching within the department and being encouraged to explore within their own teaching. An assumption implied by these results is that there can be an "inhibiting" or "supportive" departmental climate around teaching. However, no studies have operationalized these constructs. Moreover, although departmental climate has been advanced as a barrier to instructional change, few studies have explored the relationship between adoption of learner-centered instructional practices and departmental climate around teaching (Bathgate et al., 2019; Borda et al., 2020; Lund & Stains, 2015).

Several instruments have been developed to measure the climate of an organization and could thus be leveraged to design an instrument focused on departmental climate around teaching (Table 1). However, only five out of the eleven instruments in Table 1 were developed for the higher education setting, and within this sample, few focused on the department as the organization. Other instruments measure the climate of K-12 and industry organizations. Moreover, a review of the studies designing and employing these instruments pointed to methodological and analytical shortcomings (Patterson et al., 2005). First, most climate surveys lack the validity measures recommended by the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). As Table 1 indicates, 42% of the reviewed instruments did not provide evidence of content test and/or internal structure; none of the instruments included cognitive interviews to provide evidence of the type of validity known as response processes. There is thus a need to design a climate survey following the recognized standards for instrument development. Second, studies that focused on measuring climate around teaching within an organization (i.e., school, institution, or department) characterized individual faculty's description of their colleagues' perceptions of the climate within their organization, which is known as the psychological collective climate. However, the industry-focused and K–12 climate literature measures the extent to which there is a consensus among the respondents of the

Table 1 List of surveys measuring climate within an organization

Year	Name of survey	Author(s)	Type of organization	Description of instrument development	Content test	Response process	Internal structure	Relations of other variables
1963	Organizational Climate Description Questionnaire	Halpin &Croft	Elementary school	No	No	No	No	No
1967	Relationship of Centralization to Other Structural Properties	Hage & Aiken	Social welfare and health agencies	Yes	Yes	No	No	Yes
1993	Climate Survey	Ostroff	Secondary school	Yes	No	No	No	Yes
1997	The Organizational Climate Questionnaire	Furnham.& Goodstein	Industry	Yes	Yes	No	No	Yes
2005	Organizational Climate Measure	Patterson et al.	Manufacturing sector	Yes	Yes	No	Yes	Yes
2007 2012	Higher Education Research Institute Faculty survey	Lee Hurtado et al.	Higher education	Yes	No	No	Yes	Yes
2009	Organizational Change Questionnaire	Bouckenooghe, Devos & Van den Broeck	Public and private sector organizations	Yes	Yes	No	Yes	No
2012	Department Teaching Climate	Knorek	Higher education	Yes	Yes	No	Yes	Yes
2014	Survey of Climate for Instructional Improvement	Walter, Beach, Henderson & Williams	Higher education	Yes	Yes	No	Yes	Yes
2017	Instructional climate survey	Landrum, Viskupic, Shadle & Bullock	Higher education	Yes	Yes	No	Yes	No
2018	Collaborative on Academic Careers in Higher Education	Mamiseishvili, & Lee	Higher education	Yes	Yes	No	Yes	Yes

organization; this is known as the organizational collective climate (Ostroff, 1993; Schneider et al., 2013). In other words, studies need to statistically investigate whether faculty within a department are answering a climate survey in a similar way and therefore have a consensus view on the climate. Such measurement of departmental consensus is needed in order to substantiate claims regarding the importance or lack thereof of departmental collective climate on teaching in relation to the adoption of learner-centered instructional practices. No studies that we could find on departmental climate around teaching attempted to measure departmental collective climate.

This study is the first empirical exploration testing the extent to which psychological collective climate and departmental collective climate around teaching predict the level of uptake of learner-centered instructional practices. We developed and collected the Departmental Climate around Teaching (DCaT) survey from STEM faculty at 4-year institutions of higher education in the USA to explore the following research questions:

1. What *validity and reliability evidence* support the use of the Departmental Climate around Teaching (DCaT) survey?

2. To what extent do different *types of psychological collective climate around teaching* exist within STEM departments?

3. To what extent does a *departmental collective climate around teaching* exist at the departmental level?

4. To what extent does the *psychological collective climate around teaching* and *departmental collective climate around teaching* relate to the use of learner-centered instructional practices within a department?

Conceptual framework
Organizational culture versus organizational climate
Organizational culture and organizational climate are two terms employed to describe people's experiences in the settings of their working environment (Schneider et al., 2013). These terms are often misused in the literature (e.g., studies focused on culture mistakenly measure climate and vice versa), and thus, it is essential to understand the differences between these two terms.

Organizational culture is defined as "the shared basic assumptions, values, and beliefs that characterize a setting and are taught to newcomers as the proper way to think and feel, communicated by the myths and stories

people tell about how the organization came to be the way it is as it solved problems associated with external adaptation and internal integration" (Schneider et al., 2013, p.362). Organizational culture applied to the academic department represents for example the way faculty members follow unwritten rules when making departmental decisions for promotion and tenure. Culture is typically studied through qualitative case studies.

Organizational climate refers to "the shared perceptions of and the meaning attached to the policies, practices, and procedures employees experience and the behaviors they observe getting rewarded and that are supported and expected" (Schneider et al., 2013, p. 362). Organizational climate applied to the academic department represents the faculty members' current perceptions and attitudes of how teaching is evaluated. Research focused on organizational climate typically relies on surveys; data are analyzed in an aggregated way by using different statistical methods and different consensus models. For the remainder of this study, we will be focusing on the latter (i.e., organizational climate, with academic department as the organization).

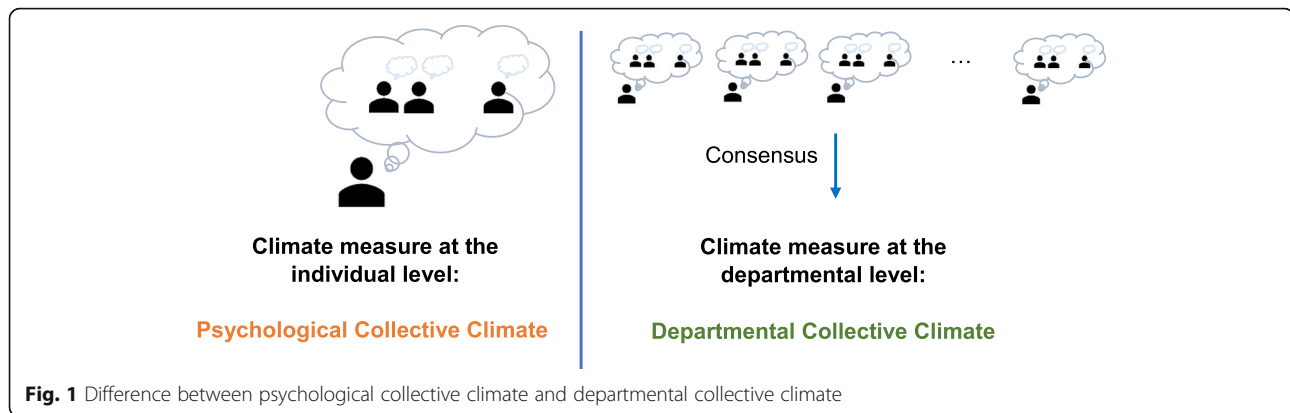### Two measures of organizational climate
Approaches to measure climate within an organization differ between studies in Discipline-Based Education Research (DBER) and studies in K–12 and industry settings. In both fields, researchers often ask each organizational member how they think others in the organization perceive organizational policies, practices, and procedures (e.g.,Ostroff, 1993; Patterson et al., 2005). However, in the K–12 and industry studies, researchers then leverage statistical methods to explore the extent to which there is a consensus on these perceptions among the members of the organization (Ehrhart & Schneider, 2016). If consensus is demonstrated, they aggregate the individual level data into a single climate measure. At this point, the measure represents the collective view of the organizational climate (i.e., the climate at the organization level (e.g., school) and not the individual level (e.g., teacher)). This aggregated measure is then used to test whether the organizational collective climate has an impact on a desired outcome for this type of organization (e.g., productivity, job satisfaction). In contrast, studies on institutional or departmental climate in DBER aggregate the perceptions collected from individual members (e.g., take an average of the perceptions across members of a department) without testing for consensus across respondents (Landrum et al., 2017; Ngai et al., 2020). The aggregated value in these studies represents the average perception of the organizational climate by members of the organization (i.e., individual level measure of climate of the institution or department) and not the collective perception of the

organizational climate (i.e., departmental or institutional level measure of climate).

The distinction in the analytical approaches employed to measure organizational climate between the K–12/industry literature and the DBER literature is important when studies are interested in exploring the relationship between organizational climate and other organizational outcomes. Although the two types of literature refer to what seems to be the same construct, organizational climate, DBER studies describe average individual perceptions of the climate while the organizational climate literature describes the collective perception of the climate. This different level in measuring organizational climate impacts claims that can be made about the relationship between such climate on organizational outcomes or characteristics such as uptake of learner-centered instructional practices.

Chan (1998) highlighted the importance of explicitly identifying the level at which the organizational climate construct is being measured. He identified five composition models to assist researchers in developing a common framework to explain the relationship between constructs and level of the organization: "Composition models specify the functional relationships among phenomena or constructs at different levels of analysis (e.g., individual level, team level, organizational level) that reference essentially the same content but that are qualitatively different at different levels" (Chan, 1998, p. 234). Making the transformation of a construct across levels of the organization clearly provides "conceptual precision in the target construct" (Chan, 1998, p. 234).

Of the five models described by Chan, the approach followed by the K–12/industry literature matches the Referent-Shift Consensus Model (Ehrhart & Schneider, 2016). Applied to this study, the measure in the Referent-Shift Consensus Model describing the climate at the department level represents the *shared* perceptions of the climate among the members of the department. This is typically done by calculating within-group agreement indexes (e.g., r_{wg}, (O'Neill, 2017)) for each factor or item in the survey and only aggregating the factors or items that meet a certain within-group agreement threshold. Moreover, the Referent-Shift Consensus Model shifts the focus from a member's personal perception of the climate in the organization to the member's thinking about how other members in the organization perceive the climate. For example, the following survey item "I embrace other colleagues' innovative teaching practices" would be rephrased "Overall, instructors in my department embrace other colleagues' innovative teaching practices." The model thus leads to two measures of climate (Fig. 1):

Fig. 1 Difference between psychological collective climate and departmental collective climate

1) "*Psychological collective climate*, is defined as the individual's description of other organizational members' perceptions of the climate" (Chan, 1998, p. 238); This measure thus focuses on the individual members of the organization and their views on how others in their organization think of the climate; this is a measure of the climate of the organization at the individual level.

2) *Organizational collective climate* is a measure of the climate at the organizational level; this measure derives from the aggregation of measures collected at the individual level (i.e., psychological collective climate); aggregation is only justified when consensus and agreement among individuals' psychological collective climate perceptions have been demonstrated. Given that the definition of organizational climate is the *shared* perceptions of policies and practices and thus a property of the organization (Ehrhart & Schneider, 2016), *organizational collective climate* is the only measure out of the two that represents that construct.

The DBER literature has been measuring the psychological collective climate. However, the analytical approach typically employed in these studies does not follow the Referent-Shift Consensus Model and rather matches a different model identified by Chan: the Additive Model. In the Additive Model, the measure at the department level is the summation or average of measures collected at the faculty level. However, the variance among answers provided by the faculty is not considered relevant during this transformation. These studies thus do not measure shared perceptions of the climate (i.e., organizational (e.g., departmental) collective climate).

At the time of writing, it is unclear whether the psychological collective climate and/or organizational collective climate is a predictor of instructional innovation within a department. In this study, we measured both the psychological collective climate and departmental collective climate through a new survey instrument that

underwent rigorous validity and reliability studies. We then tested the extent to which each measure of climate predicted the level of use of learner-centered instructional practices within the department.

Methods

The survey employed in this study consisted of three parts: (1) the Departmental Climate around Teaching (DCaT) survey, (2) an abbreviated version of the Measurement Instrument for Scientific Teaching (MIST; Durham et al., 2017), and (3) a demographic section (Table S4). All survey components are presented in Additional file 1. Participants spent an average of 20 min to answer this three-part survey, which was collected online via Qualtrics. All stages of the study were approved by the University of Nebraska-Lincoln Institutional Review Board.

Participants

To measure departmental collective climate, it is important to collect data from a large proportion of the faculty members in the department. Moreover, to address the fourth research question—relationship between climate and uptake of learner-centered instructional practices—it is necessary to capture a diverse set of instructional practices among the population surveyed. We devised two different strategies to maximize the likelihood of achieving these two goals. First, we selected STEM departments that had taken part in some department or institution-level change initiative related to teaching. Our goal was to identify settings where there may be some buy-ins in assessing climate around teaching. We read 48 abstracts of grants funded by the National Science Foundation under the WIDER program (National Science Foundation, 2013) and identified eleven projects indicating change in the departmental or institutional climate around teaching as one of their goals. We contacted the principal investigators of these projects and probed their interest in implementing our survey with their local population. One of the principal investigators

helped us identify two departments at their institution that were likely to provide a high participation rate. Second, we selected departments with faculty who had participated in national professional development programs around teaching, specifically the Cottrell Scholar Collaborative New Faculty Workshop (CSC NFW; Cottrell Scholars Collaborative, 2017) and Process-Oriented Guided Inquiry (POGIL; POGIL Team, 2019) programs. For the CSC NFW, we selected departments in which at least three of the faculty had participated in the workshop. For the POGIL programs, we contacted the program leadership to help us identify departments in which POGIL was consistently implemented by at least one of the faculty.

In total, 727 emails were sent to faculty members representing 21 different institutions. We emailed all faculty members across ranks (e.g., lecturer, tenure-track faculty) within each of the 22 departments. In total, 201 instructors completed the survey, which corresponds to a raw response rate of 28%. Raw response rates by department are provided in Table S7. The data set was then cleaned by deleting (1) participants who had no teaching responsibility, (2) participants who did not answer all items, (3) participants who used "I don't know" option for any of the items in the DCaT survey (Cole, 1987), and (4) items for which more than 5% of the participants answered "I don't know" (Cole, 1987). The cleaned sample size was 166 instructors which corresponds to a 23% response rate. Twenty-two departments are included in the sample, most of them being chemistry departments (Table S7).

Development of the Departmental Climate around Teaching (DCaT) survey

We followed guidelines from the *Standards for Educational and Psychological Testing* to develop the DCaT survey. In particular, we abided by the following steps to test the extent to which the survey provided valid and reliable data about psychological collective climate: (1) test content, (2) response processes, (3) internal structure, and (4) internal consistency.

Test content

Test content is used to test whether an instrument is capturing the intended domain. In our case, it was the psychological collective climate (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). Test content is typically established through consultation with experts in the target domain. We conducted a literature review to identify typical constructs considered by studies measuring organizational climate. This review focused on studies describing the development and/or use of surveys measuring organizational

climate and was not limited to educational settings since this body of work is minimal. Studies were selected based on whether they (1) measured climate around teaching (Knorek, 2012; Landrum et al., 2017; Walter et al., 2014), (2) measured institutional or organizational climate from an instructor's perspective (Eagan et al., 2014; Halpin & Croft, 1963; Mamiseishvili & Lee, 2018; Ostroff, 1993), or (3) described popular instruments designed to measure organizational collective climate outside of an education setting (Furnham & Goodstein, 1997; Hage & Aiken, 1967; Patterson et al., 2005; Thomas & Tymon Jr, 2009). Eleven studies met these criteria (Table 1) and were used to identify common constructs assessed across these studies as these would indicate a certain level of saliency when describing an organizational climate. We then leveraged items from these surveys to develop the first version of our survey. This initial draft (DCaT Version 1; Table S1) was then shared with DBER experts for their feedback. This feedback was used to modify the survey (DCaT Version 2; see Results section).

Response process

This step focuses on testing whether the survey items and methods used to collect answers within the survey (e.g., type of Likert scale) are interpreted by the study participants the way the developers of the survey had intended. Testing for response processes is typically conducted via cognitive interviews (Peterson et al., 2017). We gathered a convenient sample of 11 faculty members from seven institutions in the USA to participate in the cognitive interviews (Robinson, 2013). Three of the faculty members worked at primarily undergraduate institutions, while the other eight worked at research-intensive institutions. The academic appointments of the participants included two full professors, six associate professors, one assistant professor, one associate professor of practice, and one lecturer. Nine were chemists, and two were biologists. The diverse characteristics of the interview participants support the use of the survey for instructors at different academic ranks and from different types of institutions. The interviews were conducted once with each faculty in four rounds (about three faculty per round) with revisions of the survey between each round. All interviews were conducted within a 3-month period. The cognitive interviews were conducted via the Zoom online platform or in a private room to ensure the confidentiality of the participants.

Participants were first asked to fill out the DCaT Version 2 and demographic portion of the survey. Within a week following completion of the survey, we interviewed them. Interviews typically lasted 30 min and engaged the faculty members in a think-aloud process to identify their understanding of items and rationale for choosing

a response option for an item, to unpack inconsistencies within their responses, and to collect their feedback on the survey design (Table S10). The cognitive interviews led to further refinement of the survey (see Results section). This DCaT Version 3 (Table S2) was used to test the internal structure of the survey.

Internal structure

Establishing the internal structure of an instrument consists of providing evidence about the relationships between items and the constructs intended to be measured. The third version of the DCaT survey along with the adapted MIST (Table S5) and demographic section (Table S4) were embedded in Qualtrics as an online instrument. Confirmatory Factor Analysis (CFA) was employed to validate the internal structure (Arjoon et al., 2013) of the survey. Due to the categorical nature of the Likert scale responses, the weighted least square mean and variance adjusted (WLSMV) estimator was used to conduct the CFA in Mplus version 7.4. A minimum recommended sample size to conduct CFA is five to ten respondents per item (Brown, 2014). In this study, the sample size met the minimum criteria of five participants per item. Since we did not expect certain items within a construct to be more important than others in describing the construct, we followed the guidelines provided by Komperda et al. (2018) and used the Tau-equivalent indicator. In this process, all factor loadings within the same factor are set to be identical to ensure that each item within the same construct is equally weighted. This allowed us to use the mean across items within a factor to describe the factor.

The model fit was tested by exploring two typical model fit indexes for CFA analyses that are independent of sample size: the comparative fit index (CFI) indicates an adequate fit when its value is above 0.9 (Hu & Bentler, 1999) (range from 0 to 1); the root mean square error of approximation (RMSEA), with values less than 0.08 indicating a good fit (Browne & Cudeck, 1993).

Internal consistency

Testing for internal consistency provides evidence that respondents to an instrument answer similar items in similar ways. The internal consistency of the survey was measured with Cronbach's alpha coefficient since it has been suggested to be an acceptable reliability measure for Tau Equivalent models (Komperda et al., 2018). A cut-off value of > 0.7 is recommended (Komperda et al., 2018; Streiner, 2003).

Abbreviated version of the Measurement Instrument for Scientific Teaching (MIST)

The literature on instructional change indicates that faculty members often point to their departmental climate

as a barrier to using learner-centered instructional practices. The level of use of these practices was measured using an abbreviated version of the Measurement Instrument for Scientific Teaching (MIST; Table S5; Durham et al., 2017). This instrument was chosen among many others because it underwent the most rigorous validation and reliability studies (Durham et al., 2018). MIST measures the extent to which STEM faculty employ Scientific Teaching practices in their classroom through self-report from faculty. The instrument measures eight sub-categories of Scientific Teaching practices (Active-learning strategies, learning goals use and feedback, inclusivity, responsiveness to students, experimental design and communication, data analysis and interpretation, cognitive skills, and reflection). While all these categories fall under learner-centered instructional practices, in this study, we used only the Active-learning strategies subcategory since it has been demonstrated to be the MIST subcategory that correlates the strongest with students' and external observers' reports of the level of active learning in a course (Durham et al., 2018). Considering our expected modest sample size to establish internal structure, we consulted with the MIST developers to identify a subset of five items within that Active-learning strategies subcategory that statistically correlated best with external measures of presence of active learning. These five items include (1) average percent of class time during which students were asked to answer questions, solve problems, or complete activities other than listening to a lecture; (2) frequency of use of polling methods; (3) frequency of use of in-class activities other than polling methods; and frequency of (4) in- and (5) out-of-class group work activities.

Confirmatory factor analysis for this abbreviated version of MIST was conducted using Mplus version 7.4 with a Robust Maximum Likelihood (MLR) estimator and congeneric indicator. The comparative fit index (CFI) for all five items was 0.981, and the standardized root mean square residual (SRMR) value was 0.039. However, the factor loadings for two items (frequency of use of polling methods and frequency of group work for outside the lecture hall) were lower than 0.2 (Sharma et al., 2005) and thus deleted. Since there were only three items left to describe this one factor, a large sample size was needed (more than 300) to conduct the CFA (MacCallum et al., 1999). Our sample size (N = 166) did not meet this criterion, so we relied on a reliability test. Cronbach's alpha for the abbreviated version of MIST was 0.80, which indicated that these three items were answered in a consistent manner by participants.

A MIST scale score was generated following the equation given in the original paper for each participant (Durham et al., 2017):

$$\text{MIST}_{\text{scale score}} = [(XQ1 + XQ2 + ... + XQn)/n] \quad (1)$$

where XQ1... XQn are the normalized response for each question, and n is the number of items included in the scale score calculation. The MIST score was used as a dependent variable in a simple linear regression analysis.

Analysis

Mixture model cluster analysis

Cluster analysis was used to identify groups of faculty members who provided similar psychological collective climate descriptions. Mixture model clustering (MMC) analysis was used in this research for the following reasons:

1) The MMC method, unlike the heuristics methods of clustering such as K-means is able to make population-wide estimations (Landau & Chis Ster, 2010);
2) The heuristic methods require a preassigned number of clusters before the analysis, while the MMC method provides empirical recommendations for the number of clusters (Fraley & Raftery, 1998);
3) The MMC method allows for comparison of different solutions by using fit indices.

Bayesian Information Criterion (BIC) was used to identify the class enumeration.

Regression analysis to explore the relationship between climate and instructors' uptake of learner-centered instructional practices

A simple linear regression was used to predict instructors' use of learner-centered instructional practices based on types of psychological collective climate. Since the type of psychological collective climate was categorical in nature, dummy coding was used to recode this variable.

$$y_i = \beta_0 + \beta_1 X_1 + ... \beta_i X_i \quad (2)$$

y_i is the dependent variable which refers to the abbreviated MIST scores in this study, X through X_i are different clusters identified in the MMC analysis which served as independent variables, β_0 is the constant in the regression equation, and β_1 through β_i are standardized coefficients for different clusters.

Measuring the departmental collective climate

The measure of departmental collective climate necessitates a high response rate. This measure was calculated for the three departments that initially had at least half of their instructors answer the DCaT survey (see raw

response rate in Table S7). The cleaned response rates for these three departments are provided in Table S8.

Inter-rater agreement, $r_{wg(J)}$, which indicates the level of agreement among faculty members within the same department, was calculated using the following equation:

$$r_{wg(J)} = \frac{J\left(1 - \frac{\bar{s}_x^2}{\bar{s}_{EU}^2}\right)}{J\left(1 - \frac{\bar{s}_x^2}{\bar{s}_{EU}^2}\right) + \frac{\bar{s}_x^2}{\bar{s}_{EU}^2}} \quad (3)$$

where, J is the number of items in the construct, $\bar{s}_x^2$ is the obtained average variance of the items in the construct, and $\bar{s}_{EU}^2$ is the variance of the uniform distribution. A value of $r_{wg(J)}$ above 0.75 indicates a high level of within-group inter-rater agreement (O'Neill, 2017).

Intra-class correlation coefficients (ICC) do not investigate absolute agreement but consistency between faculty members within a department. Two coefficients were calculated: ICC(1) and ICC(2). ICC(1) indicates whether the mean of one faculty is a reliable measure of the mean of another faculty within the same department. ICC(2) indicates whether there is a reliable difference between means across different departments (Koo & Li, 2016). The equations for calculating ICC(1) and ICC(2) are listed below:

$$\text{ICC}(1) = \frac{\text{BMS} - \text{WMS}}{\text{BMS} + (k-1)\text{WMS}} \quad (4)$$

$$\text{ICC}(2) = \frac{\text{BMS} - \text{WMS}}{\text{BMS}} \quad (5)$$

where BMS is the between-treatments mean square, and WMS is the within-treatment mean square.

Results

Validity and reliability evidence supporting the use of the Departmental Climate around Teaching survey

Modifications based on the test content study

Constructs and items were initially identified through a literature review of studies measuring teaching climate or organizational climate within an educational setting as well as highly cited studies describing a survey measuring organizational collective climate in organizations outside academia. Twelve studies were identified through this review process (Table S6 contains a list of the studies and the constructs measured in each). We considered constructs that had been included in at least four of these studies, which included Formalization, Co-operation, Participation, Supervisor Support, Warmth, Growth, Innovation, Autonomy, Achievement, Extrinsic Reward, Performance Feedback, and Outward Focus. One of these constructs was not included in the first version of the DCaT survey: Formalization—i.e.,

"perception of formality and constraints in the organization; emphasis on rules, regulations and procedures" (Ostroff, 1993, p. 62)—was a construct found in K-12 and industry-focused studies and was not considered relevant when exploring the climate around teaching in the higher education setting. We added the construct of Resources since funding, space, and teaching budget have been identified as barriers to uptake of learner-centered instructional practices in prior studies (Sturtevant & Wheeler, 2019) and this construct was included in the most recent climate surveys developed for STEM higher education contexts (Landrum et al., 2017; Walter et al., 2014). The first draft of the survey (DCaT Version 1, Table S1), which contained 48 items was developed based on this analysis. Thirty six items came from these prior studies (as is or modified if needed to adapt to context), and 12 items were developed by the authors. This draft was shared with the University of Nebraska-Lincoln DBER group during Spring 2018 to collect their feedback. The survey was further revised as a result of this consultation (DCaT Version 2). For example, the DBER group pointed out that the targeted population was not clearly described. Throughout the survey, we had used the word "instructors", but it was unclear to the DBER group who should be included as an instructor. For example, certain STEM departments heavily rely on graduate teaching assistants to teach lectures, laboratories, or recitations. To clarify our targeted population, we added the following description at the beginning of the survey: "Instructors" in this questionnaire refer to faculty who teach undergraduate level courses (including lecturer, tenured/tenure-track professor, professor of practice, but EXCLUDING graduate students). Graduate teaching assistants were not included as instructors since they have limited involvement in decisions related to teaching and curriculum conversations within a department (e.g., they are not assigned to or made aware of decisions made during curriculum committees).

Modifications based on the response process study

The revised version of the survey (DCaT Version 2) was then distributed to faculty members who had agreed to take part in cognitive interviews. Faculty answered the survey prior to being interviewed (see the Method section for details on the response process procedures). The cognitive interviews helped us to identify several critical issues with this version of the instrument.

The first issue was that the participants did not consider the departmental climate as a shared perception of all instructors within their department when they answered the questions. Items in several of the surveys identified in the literature started with "Instructors in my department". We either used these items or

developed items that used the same wording. However, interviewees highlighted that this could lead to different interpretations as the following quote exemplifies: "*When answering questions, you use frequency, others may use your own data points. Some items, you may think about a certain meeting*" (Participant #4). We thus changed each item by adding at the beginning the word "overall". This change was made halfway through the collection of the interviews. Participants interviewed afterwards demonstrated an understanding that instructors should be considered as a collective when answering the survey.

Second, it became clear during the interviews that the response option provided—a 5-point Likert scale going from strongly agree to strongly disagree—did not clearly capture participants' opinions, especially when choosing the neutral option (i.e., neither agree nor disagree): "*I can't strongly answer that. Some of them [items], like I wonder if you ... so you don't have an N/A or can't answer category, right? Which, so I kind of was using that middle one [neither agree nor disagree] as that*" (Participant #4). We had debated as to whether the neutral position should be included in the first place since many of the surveys we were consulting did not include it. However, the interviews confirmed to us that it was needed and that another option "I don't know" should be added. Including these two extra response options helped respondents feel that they were not forced to agree or disagree when they were truly on the fence or did not have enough information to decide. It also helped us ensure that the neutral option was interpreted consistently across the participants.

Third, most of the climate surveys do not ask the participants to answer the survey based on a particular time frame. This issue was raised during the interviews as the following quote illustrates: "*Some questions need a time frame. Past five years? Or past three years? The answer will be different*" (Participants #9). For example, a faculty member who has been in their department for 20 years may have experienced different climates during their time and may try to think across all of those when responding to the survey. Since prior studies have shown that departmental leaderships can influence instructors' perceptions of the teaching context within the department (Ramsden et al., 2007), we decided to align the time frame with the period of activity of the current department chair. We thus added the following question prior to answering the DCaT survey (Table S4): "How long has your chair been in his/her current position?" We then added the following sentence to the instruction for the DCaT survey, leveraging the 'piped text' option in Qualtrics (Table S3): "When answering the survey, please focus on the last '*piped text from the chair question*' year(s)". We purposefully did not refer to the chair

in the instruction so as to limit potential (conscious or unconscious) influence that it could have on respondents.

Fourth, several constructs typically included in climate surveys were eliminated as a result of the analysis of the cognitive interviews: Extrinsic Rewards, Resources, and Warmth. Interviewees indicated having trouble answering items associated to the Extrinsic Rewards construct for several reasons. One of the faculty from a primary undergraduate institution indicated that the evaluation of teaching did not occur at the department level but rather at the provost level. Moreover, hiring processes at their institution were conducted at the college level not the department. Therefore, reward for teaching excellence in the form of hiring or promotion was not controlled by the department, making the extrinsic award items irrelevant for this participant. Another participant from a primary undergraduate institution indicated that evaluation of teaching was conducted at the Chair and Dean's levels. We assume that these situations may not be unique among primary undergraduate institutions. One of the items for extrinsic rewards focused on teaching awards (see Table S1). One of the participants at a research-intensive institution indicated that factors other than teaching excellence may come into play for teaching award nominations:

> "People are nominated for awards a lot, but that doesn't mean that you need to be a really great teacher to continue in your job. [...] The department nominates for teaching awards every year because we're not just going to sit out on a cycle where we're going to nominate people for awards no matter what." Participant #1

Moreover, four participants indicated that there are no respected, highly desirable awards for teaching like there are for research.

> "So you don't get pay raises for your teaching, don't get recruited. You don't get counteroffers. The administration, they pass around some awards and what are those awards giving you? There's $0 million behind that award." Participant #4

> "We know that the money follows the research awards, but doesn't follow the teaching awards." Participant #5

Similarly, four of the eleven interview participants indicated that there was no good measure of effective teaching, making it difficult to reward and recognize it.

> "We place a high premium on quality, but we have no metrics to assess that. So if one, if a chair think

you're a good teacher, then you're a good teacher" Participant #3

> "We don't have any sort of standardized measure of teaching performance. [...] There are actually very few sort of strong public indicators of teaching achievements." Participant #8

Consequently, participants found it challenging to answer the item "Overall instructors in my department carefully consider evidence of effective teaching when making decisions about continued employment and/or promotion".

The Resources construct captures the tools used for instructional improvement, including time, funding, office space, equipment, and support services. Our interviewees indicated that most of these resources were not controlled by the department and thus did not feel it was appropriate to ask these questions: "*Time as a resource is not controlled by the department*" and "*Resource is provided at the college level, not at the departmental level*". The Warmth construct, which captures whether informal communications occur among faculty members, was eliminated because faculty members indicated that they did not have enough information to answer this type of questions. Here are some quotes to illustrate this point:

> "That's actually really hard to get these questions, unless you know everyone very well informally and attend these happy hours or lunches and things like that, which I do not know if many, I don't know the answer." Participant #9

> "Oh yeah. Yeah. I have no clue. So in that case, I don't know what my colleagues are doing and so I do not know." Participant #10

Fifth, the Participation and Cooperation constructs were combined. The construct Participation was defined as measuring whether faculty members were involved in decision-making process and setting goals/policies with respect to the teaching mission of the department. Two of the items were removed since the cognitive interviews illustrated that adoption of teaching methods does not occur at the departmental level: "*I don't know if maybe this, all right, so we don't as a department make decisions on new teaching methods*" (Participants #4). Regarding the construct Cooperation, two of the original items were moved to "supervisor support" which is more focused on capturing the helpfulness of the departmental leadership. Since both constructs Participation and Cooperation had only two items each and were conceptually similar, we combined them into one construct,

Involvement. We measured the internal consistency of the four items by using Cronbach's alpha, and obtained a 0.85 indicating that the four items were measuring similar ideas. We named the new construct Involvement which we defined as faculty members' perceptions of involvement in decision-making processes, setting goals, policies with respect to the teaching mission of the department, and the willingness to communicate about teaching related issue with colleagues.

Following the cognitive interviews, the third version of the DCaT survey (Table S2) consisted of eight constructs (32 items): Involvement, Growth, Autonomy, Supervisor Support, Innovation, Outward Focus, Achievement, and Performance Feedback. Each construct's definition is provided in Table 2. A six-point Likert scale was used as the item response options from "strongly disagree" (as 1) to "strongly agree" (as 5) with the sixth point represented by "I don't know".

Internal structure of the DCaT survey

Before evaluating the internal structure of the DCaT survey, the data set was cleaned up (see Methods section). Items for which more than 5% of the participants answered "I don't know" were deleted (supervisor support-1 and achievement-3). Consequently, 30 items were used to explore the internal structure of the DCaT survey.

The comparative fit index (CFI) for all 30 items was 0.654 (below the 0.9 threshold) indicating a lack of fit. We consulted the Modification Indices (M.I.) in order to identify items that could improve the model fit (Ab Hamid et al., 2011). We considered whether to eliminate items with high M.I. values based on their alignment with the construct they were intended to measure. As a result, the item performance feedback-4 was deleted. Indeed, this item had high M.I. value and was focused on the measures used to evaluate teaching rather than the feedback provided to faculty and was thus not as well aligned with the construct as the other three items. Another CFA analysis was conducted on the 29 items

(Table 3). The CFI for this model was 0.946 (above the 0.9 threshold) and the root mean square error of approximation (RMSEA) was 0.076 (below the 0.080 threshold). These results thus indicated that this final version of the DCaT survey (DCaT Version 4; Table S3) met the goodness-of-fit standards.

Internal consistency

Internal consistency of each construct was evaluated with Cronbach's alpha (Table 4). We observed values ranging from 0.77 to 0.93, which indicated that the data from the DCaT Version 4 were sufficiently reliable for interpretation.

Types of psychological collective climate around teaching in STEM departments

Once it was demonstrated that the DCaT survey provided valid and reliable data, we endeavored to answer the second research question: To what extent do different types of psychological collective climate around teaching exist within STEM departments?

Overall, faculty members in this study reported a neutral to positive psychological collective climate around teaching in their department as indicated by the range of means for each construct (from 2.87 ± 1.00 to 3.95 ± 0.73 on a scale from 1 to 5, 5 being strongly agree; Table 4). We conducted a Mixture Model Clustering analysis to identify groups of faculty members across the whole sample who provided similar psychological collective climate descriptions. To identify the model that best fit the data, we examined the Bayesian Information Criterion (BIC) and the BIC difference (Table 5). Although the "VEE" model had the lowest BIC value, the sample within this model was evenly split between the two clusters and the interpretation of the clusters lacked nuance (i.e., estimated means were similar across the two clusters for most of the constructs) (Table S9). The next best model, "VEI", had a four-cluster solution, which provided an interpretable and nuanced description of the

Table 2 Definitions of the constructs measured in the Departmental Climate around Teaching survey

Construct	Definition
Involvement	Perception that faculty are involved in decision-making processes and setting goals and policies with respect to the teaching mission of the department.
Growth	Perception of emphasis on personal development with respect to teaching
Supervisor support	The extent to which instructors experience support and understanding from their department chair
Autonomy	Perception of having the freedom to plan and control over their own curriculum
Innovation	Perception of emphasis on innovation and creativity in teaching approaches and accepting of changes in instructional practices
Outward focus	The extent to which the department is responsive to the needs of students
Achievement	Perception of challenges, demand for work, and continuous improvement of performance
Performance feedback	The measurement of and feedback on teaching performance

Table 3 Standardized factor loadings from the confirmatory factor analysis of the DCaT survey ($n = 166$)

Factor	Item	Factor loading
Involvement	Invo1	0.84
	Invo2	0.84
	Invo3	0.84
	Invo4	0.84
Growth	Grow1	0.79
	Grow2	0.79
	Grow3	0.79
	Grow4	0.79
Supervisor support	Supp2	0.88
	Supp3	0.88
	Supp4	0.88
Autonomy	Auto1	0.84
	Auto2	0.84
	Auto3	0.84
	Auto4	0.84
Innovation	Inno1	0.85
	Inno2	0.85
	Inno3	0.85
	Inno4	0.85
Outward focus	Outw1	0.77
	Outw2	0.77
	Outw3	0.77
	Outw4	0.77
Achievement	Achi1	0.88
	Achi2	0.88
	Achi4	0.88
Performance feedback	Feed1	0.93
	Feed2	0.93
	Feed3	0.93

Table 4 Cronbach's alphas, means, and standard deviations for each construct measured in the DCaT survey ($n = 166$)

Constructs	Cronbach's α	Mean	SD
Involvement	0.83	3.75	0.85
Growth	0.78	3.11	0.80
Autonomy	0.82	3.95	0.78
Supervisor support	0.87	3.67	0.94
Innovation	0.88	3.43	0.78
Outward focus	0.82	3.56	0.84
Achievement	0.85	3.95	0.73
Performance feedback	0.83	2.87	1.00

Table 5 Mixture Model Clustering (MMC) class enumeration ($n = 166$)

Model type, number of class	VEE, 2	VEI, 4	VEI, 3
BIC	− 2930.7	− 2945.5	− 2946.4
BIC difference	0.0	− 14.8	− 15.7

types of climate perceived by the sample. Consequently, VEI was selected for the rest of the analysis.

The estimated model proportion of each of the four types of climate are listed in Table 6 along with the estimated means for each construct. We leveraged results presented in Table 6 to describe each climate. Of note, the construct Autonomy had limited variation across all four types of climate; mostly, participants felt that faculty members in their department had a lot of autonomy with respect to teaching. As the color code provided in Table 6 indicates, we see an evolution between climates 1 to 4 from a negative to a positive description of the psychological collective climate around teaching. Thirteen percent of the participants fell into the first climate, which we label *Negative psychological collective climate around teaching*. Participants in this cluster indicated disagreeing with most of the constructs. In particular, they felt a lack of emphasis on personal development with respect to teaching ($M_{Growth} = 1.9$) and a limited ability to get feedback on their teaching ($M_{Performance\ feedback} = 1.7$). Over a third of the participants (38%) fell into the second climate, which we label *Slightly positive psychological collective climate around teaching*. Except for Performance Feedback, the mean across the constructs ranged from 2.9 to 3.7, indicating neutral to positive assessment of these constructs. A third of the participants (33%) belong to the third climate, which we label *Positive psychological collective climate around teaching*, with construct means ranging from 3.3 to 4.2. Finally, the last climate accounts for 16% of the participants. We label this fourth climate *Very positive psychological collective climate around teaching* since the construct means within this climate type are all above 4.0. Participants in this climate type reported that faculty members within their department were extremely involved with teaching-related decisions ($M_{Involvement} = 4.5$), felt strongly supported by their chair ($M_{Supervisor\ support} = 4.6$), and had a desire to excel in their teaching ($M_{Achievement} = 4.7$). We explored through Fisher's exact tests whether academic rank and tenure status were related to the type of climate and found no statistically significant relationship for either (see figures S2 and S3).

While we were able to identify different types of psychological collective climate around teaching among our diverse population of faculty members, it is unclear whether a *departmental collective climate* around teaching exists within the departments surveyed. In the next

Table 6 Estimated model proportions and estimated means for the VEI model (*n* = 166). The color coding in the table represents the average level of agreement on items within the construct, red indicating strong disagreement and green indicating strong agreement

Constructs	Type of climate (Estimated model proportion)			
	Negative (13%)	Slightly positive (38%)	Positive (33%)	Very Positive (16%)
Involvement	2.4	3.7	4.0	4.5
Growth	1.9	2.9	3.3	4.1
Autonomy	4.2	3.7	4.1	4.1
Supervisor Support	2.6	3.3	4.1	4.6
Innovation	2.3	3.3	3.5	4.4
Outward Focus	2.2	3.3	3.9	4.4
Achievement	3.1	3.7	4.2	4.7
Performance feedback	1.7	2.4	3.3	4.1

section, we explore our third research question by leveraging data collected from the three departments that provided the highest response rate.

Departmental collective climate around teaching

Three departments coded as Department 16, Department 17, and Department 22 were chosen for this analysis since they had the highest response rates across our sample (90%, 77%, and 53% respectively before data cleaning; 71%, 46%, and 44% respectively after data cleaning; see Tables S7 and S8). Although we were unable to have every single faculty member in these departments answer the survey, those that answered provide an adequate representation by academic rank (e.g.,

assistant, associate professor) of the department's composition (see Figure S1).

In Fig. 2, we provide a description of the variety of psychological collective climate present in each of these three departments. All three departments had a different distribution across the four types of psychological collective climate. All faculty members in Department 17 held a positive view towards the psychological collective climate around teaching in their department, while 13% of the faculty members in Departments 22 and 16 had negative views about their departments.

As indicated in the theoretical framework, the psychological collective climate measures a faculty member's assessment of other faculty members' perceptions of the teaching climate within the department. This measure
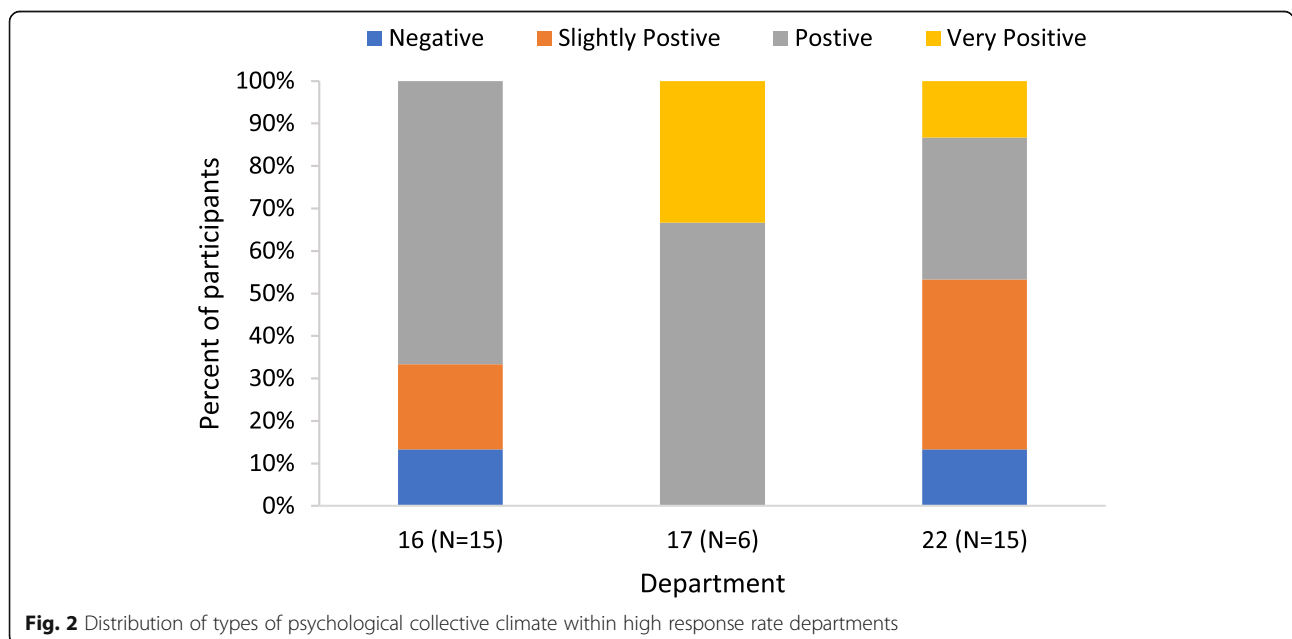

Fig. 2 Distribution of types of psychological collective climate within high response rate departments

helps us understand the different points of view within a department but does not describe the climate around teaching of the department as a whole (i.e., the shared values). To obtain the latter, *departmental collective climate around teaching*, the level of consensus on the psychological collective climate among faculty members within the same department need to be tested. This is typically assessed in the literature by calculating the inter-rater agreement index $r_{wg(J)}$ and intraclass correlation indices ICC(1) and ICC(2). $r_{wg(J)}$ indicates the level of agreement among faculty members within the same department with values below 0.75 indicating low agreement (O'Neill, 2017). As Table 7 indicates, although there was consensus within each department on most constructs, none of the departments had consensus on all constructs. Intraclass correlation indices do not investigate absolute agreement but consistency between faculty members within a department. ICC(1) indicates whether the mean of one faculty is a reliable measure of the mean of another faculty within the same department. ICC(2) indicates whether there is a reliable difference between means across different departments. Results for both of these indices are presented in Table 8, along with associated cut-off interpretations. These results indicate that only Outward focus met the inter-reliability criteria. Considering all three indices together, the only construct that could be aggregated was Outward Focus for this particular data set. Since only one of the eight constructs considered could provide the desired level of reliability, we conclude that departmental collective climate around teaching could not be measured for our three departments.

Relationships between psychological collective climate, departmental collective climate, and instructors' use of learner-centered instructional practices

The last research question focused on characterizing the relationship between the two different types of climate measure (i.e., psychological collective climate and departmental collective climate around teaching) and the level of use of learner-centered instructional practices. The abbreviated MIST instrument was used to measure the latter. Since we could not reliably measure departmental collective climate around teaching, we could not explore its relationship to instructional practices.

A simple linear regression was employed to predict instructors' use of learner-centered instructional practices

based on the types of psychological collective climate (Eq. 2). Since some participants did not answer the abbreviated MIST portion of the survey, the number of participants eligible for this analysis was 149. A non-significant regression—$F(3,145) = 1.029$, $p = 0.382$ with an $R^2 = 0.021$—indicated that in our data set, faculty members' view of psychological collective climate around teaching within their department could not be used to predict instructors' use of leaner-centered instructional practices. We went back to the cognitive interviews to identify whether faculty members provided some information that could point to a preliminary explanation for this lack of relationship. Several themes emerged that relate to a lack of communication about teaching among faculty members and lack of teaching-related standards. First, six of the eleven interviewees indicated that teaching was an independent endeavor in their departments:

> "So, so inside the room I can, [...] my students would be doing activities or whatever, but I know somebody else would just be, it'd be 50 minutes of lecture. So like we, and nobody would say anything about either of us. We were just doing our own thing." Participants #3

> "I think people tend to want to solve their own problems and figure things out for themselves. Everybody's fairly independent." Participants #7

> "I think the idea is giving people a lot of autonomy and I guess maybe we're, maybe we're sort of airing too far in the direction of autonomy and not enough on working together to think about best teaching practices, I think. I think it's just that it's kind of the culture in our department to give people as much autonomy as possible in terms of how they teach and what they choose to do." Participants #7

The survey results also clearly demonstrated the high level of autonomy that faculty members have with respect to teaching (Table 4). Second, six of the eleven participants indicated that there were no expectations for someone to improve their teaching:

> "I don't know that there is any sort of top-down expectations that that is going to be happening with

Table 7 $r_{wg(J)}$ results for high-response-rate departments. $r_{wg(J)}$ values below the 0.75 threshold are bolded

Department	Involvement	Growth	Autonomy	Supervisor support	Innovation	Outward focus	Achievement	Performance feedback
16	**0.114**	0.779	0.886	0.810	0.845	0.771	0.790	**0.400**
17	0.872	0.805	0.876	0.928	0.775	0.954	0.949	**0.618**
22	**0.735**	0.853	0.938	0.897	0.868	0.871	0.928	**0.734**

Table 8 ICC values. ICC(1) value interpretation: > 0.05 small to medium; > 0.25 large effect; ICC(2) value interpretation: < 0.4 poor; 0.40–0.75 fair to good; > 0.75 excellent. Values that met the highest cut-off are bolded

Construct	ICC(1)	ICC(2)
Involvement	0.067	0.410
Growth	0.139	0.610
Autonomy	− 0.088	− 3.595
Supervisor support	0.017	0.140
Innovation	0.005	0.043
Outward focus	**0.268**	**0.781**
Achievement	0.087	0.480
Performance feedback	− 0.041	− 0.611

any specific frequency. Um, or that you have to demonstrate on any sort of a regular basis that you have gotten better over some sort of times." Participants #8

"I guess people are expected to get better, but like it doesn't matter if you don't either." Participant #1

Along the same vein, five indicated that there was no consensus or standard guidelines for the teaching approach one should employ:

"I don't think that as a department we make decisions about teaching methods." Participant #8

"Did we set high standards, you know, on some level, but like nobody could tell you what that standard is." Participant #3

"Like when I became an instructor, no one said, you know, you have to use a specific teaching method, right? Like, no one gave me any direction about that. [...] I don't even think that the department, you know, has guidelines for how to teach." Participant #11

Discussion

This study provides insight into the challenges in measuring climate around teaching at the departmental level and highlights the need for a more rigorous approach to measuring this construct when exploring relationships between climate and other characteristics of the department such as the uptake of learner-centered instructional practices.

Cognitive interviews helped identify issues that need to be considered when measuring teaching climate

The cognitive interviews revealed challenges with constructs often measured in surveys on organization climate. For example, the construct Resources, which was included in two of the latest STEM-focused climate surveys (Landrum et al., 2017; Walter et al., 2014) had to be removed from the survey since interviewees indicated that resources (e.g., teaching and learning assistants, budgets) are typically not decided at the department level. Similarly, the construct of External Rewards was included in four of the twelve surveys listed in Table 1. The literature points to the need to include this construct as poor reward policies for teaching are often described as barriers to adoption of learner-centered instructional practices (Sturtevant & Wheeler, 2019). In particular, faculty members in these studies typically identify a lack of reward for teaching or a heavier emphasis on research during evaluation processes. The analysis of the cognitive interviews aligned with these findings. However, the interviewees also helped us realize that rewards and evaluation of teaching are not always decided and controlled at the department level, especially at 4-year institutions. Moreover, the interviewees pointed to a lack of definition and tools to measure effective teaching. This is a common weakness of the higher education system in the USA, and several initiatives are attempting to address it (Debad, 2020; National Science Foundation, 2020). Without rigorous and validated means to measure teaching, it is challenging to reward it. Therefore, the development of these tools is paramount for extrinsic rewards to be meaningfully included as a construct when measuring teaching climate at the institutional level and for extrinsic rewards to effectively impact the climate around teaching within a higher education institution.

Finally, the cognitive interviews highlighted the need to provide a timeframe for the survey participants to consider. The literature has highlighted that contrary to culture, climate can change over a shorter time scale (Schneider et al., 2013). Moreover, studies have shown that academic leaders such as chairs and deans can influence climates within their unit (Kezar, 2016; Ramsden et al., 2007). For example, a senior faculty with 10+ years in one department may have experienced different chairs and thus may respond to the climate survey based on their overall experience across these chairs or based only on the most recent chair. Junior faculty on the other end may only have experienced one chair. Therefore, providing a timeframe that is common to all survey participants would enhance the reliability and validity of the data. One caveat with focusing on just one chair is that senior faculty's assessment of the climate under that chair would be relative to other chairs they experienced in the past. It is thus critical to have broad representation of faculty across ranks when measuring the departmental climate.

It is important to define the level of the organization targeted when measuring climate

In this study, we leveraged the organization climate literature (e.g. studies in K–12 and industry) to investigate two different measures of climate based on the level of the organization that we were interested in: psychological collective climate for the individual level and departmental collective climate for the department. This contrasts with approaches in prior DBER studies, which measured psychological collective climate but treated it as a measure of departmental collective climate. We demonstrated that different types of psychological collective climate exist within a department, further reinforcing the assertion that this measure does not represent the departmental collective climate. Moreover, only one construct in our study met within-group consensus criteria required to measure departmental collective climate. These results thus demonstrate the challenges in measuring departmental collective climate in STEM departments.

In this study, we also leveraged the K–12 and industry literature on organizational climate to identify the set of constructs that defines organizational climate. However, the cognitive interviews indicated that some of these constructs (e.g., Resources and Extrinsic rewards) are not relevant for a climate measure at the department level but would be meaningful to integrate in a climate measure at the college or institution level. Consequently, the relevancy of constructs to be included in a climate measure is dependent on the level of the organization targeted and should be explored during the initial stages of the development of a climate instrument.

Overall, future studies should carefully align the goals of their investigation and their measure of climate.

The link between climate around teaching within a department and faculty members' use of learner-centered instructional practices is more unclear than previously thought

One of the main goals of this study was to investigate the relationship between the uptake of learner-centered instructional practices and departmental climate around teaching measured at two different levels (i.e., the individual faculty and the department as a whole). We found a lack of relationship at the individual level. This is counter to numerous studies in which faculty members surveyed or interviewed had identified elements of departmental climate as barriers to instructional innovation (e.g., Henderson & Dancy, 2007; Landrum et al., 2017; Sturtevant & Wheeler, 2019). These studies had led to the assumption that a departmental climate around teaching could inhibit or support the adoption of learner-centered instructional practices. In our study, we do not find evidence to support this assumption.

Although exploring the reasons behind these findings was beyond the scope of this study, data collected here identify a potential reason: teaching is an autonomous endeavor with unclear expectations. First, we see in the survey responses and in the cognitive interviews that faculty members have a high level of autonomy when it comes to instructional strategies employed in their courses. These levels resonate with findings in previous studies (Landrum et al., 2017; Walter et al., 2014). Second, the interviews pointed to a lack of benchmark for effective teaching at the departmental level and the idea that instructional style is not imposed on faculty unless the evaluation process (e.g., student evaluation) identifies serious problems. Finally, the interviewees pointed to a lack of regular and frequent communication around teaching among departmental colleagues (as illustrated by the participants' inability to answer items related to informal conversations around teaching, Warmth construct). There is thus no shared understanding of effective teaching at the department level and limited opportunities to share best practices.

The definition of organizational climate, "the shared perceptions of and the meaning attached to the policies, practices, and procedures employees experience and the behaviors they observe getting rewarded and that are supported and expected," applied to teaching implies that there are policies around teaching and that effective teaching is rewarded. However, as we have seen in this study, effective teaching is ill-defined in most of the departments of our interviewees and consequently not measured; this results in teaching not being rewarded. Results in this study thus indicate that faculty within STEM departments have an underdeveloped departmental climate around teaching both at the individual and departmental levels. However, future studies should delve more on this topic by addressing some of the limitations of the present study.

One implication of this finding is the necessity for higher education researchers to provide evidence-based descriptions as well as valid and reliable measures of effective teaching. Participants in this study echoed what has been reported in other studies (i.e., there is a lack of rewards and recognition of teaching). It may be that the department would reward teaching effectiveness if they felt they had the right tools to measure it. Having a set of criteria, albeit probably imperfect, for effective teaching and ways to measure it would also enable the development and value of high-profile teaching-related awards similar to those that exist for research (which are also based on an imperfect set of criteria). Moreover, if awards of the same stature as research awards existed for teaching, it may be that faculty members would consider them. In other words, the lack of focus on rewarding teaching may not be due to an unfavorable climate

to rewarding teaching but rather the inexistence of high-profile ways to recognize it.

In turn, if departments and institutions leveraged this work to develop an evidence-based definition of effective teaching that aligns with their context, the conversations alone that would be required to achieve this vision would be a tremendous departure from what we have seen in this data set in terms of communication and involvement of faculty in teaching-related issues. It would help establish some of the elements of an organizational collective climate around teaching. Of course, if the adoption of such definition could lead to actionable policies that are understood, valued, and enforced by community members, then we could be in a position where organizational collective climate can actually be measured.

Limitations

Several limitations should be considered for this study. First, our sample size, even if it met the criteria for conducting CFA analysis (at least five participants per item), is small ($n = 166$). Moreover, the response rates at the department level varied widely. The response rates of the three departments that were used to measure departmental collective climate were the highest in our sample but lower than what would be desired (> 80%). These limitations associated with the sample size thus minimize the generalizability of the results. Second, the sample is biased toward research intensive institutions and chemistry departments. Eighteen of the twenty-two departments included in this study are embedded in research-intensive institutions, nineteen are chemistry departments. The results should thus be viewed in that light. However, we included faculty members from primarily undergraduate institutions as part of the development and validation process of the DCaT survey and thus believe that the DCaT survey is well suited to be implemented in these environments. Third, our sampling strategy may have skewed the results toward more positive descriptions of climate since several of the faculty members within these departments had engaged in pedagogical professional development programs. This may have increased our ability to establish the validity of the data collected by the DCaT. Finally, we chose a Likert-type scale for this survey as we were interested in measuring unobservable characteristics of the faculty (i.e., their perceptions of their colleagues' perceptions of their departmental climate), and this option format is well-suited for large-scale explorations of this type of characteristics (Ho, 2017). However, this type of scale is prone to biases such as central tendency bias, acquiescence bias, and social desirability bias (Subedi, 2016). Moreover, the 5-point scale provides low level of variations when compared with a quantitative scale.

Use of the DCaT survey

We have demonstrated that the DCaT survey can measure psychological collective climate but it is still unclear whether departmental collective climate is measurable with this instrument or other climate instruments. Caution should thus be taken when one implements this survey. If the intent is, for example, to monitor changes of the climate around teaching in a department as a reform effort is implemented, the DCaT survey can assist by evaluating changes in the distribution of types of psychological collective climate over time.

Evidence collected in this study were based on a small and bias sample. Implementation of the DCaT survey in a broader and larger sample would help further explore whether departmental collective climate is indeed not measurable. It would also be interesting to collect the DCaT survey along with a measure of climate around teaching at the college or institution level to explore overlaps between these two constructs and their differentiated abilities to explain uptake of learner-centered instructional practices. These studies as well as any studies employing the DCaT survey for research should conduct validity and reliability checks (especially cognitive interviews).

Whether the DCaT survey is used by a single department to understand their own climate around teaching or as part of a research project, we recommend triangulating the DCaT results with interviews in order to better understand the climate itself and factors influencing it. The interviews could explore in more depth some of the constructs measured in the DCaT survey; it could also probe other aspects of the climate not captured by the DCaT survey such as the role of extrinsic as well as intrinsic rewards regarding teaching; finally, the interviews could also shine some light on the influence of factors external to the department in shaping the climate around teaching (e.g., policies at the institutional and/or college level, and role of professional organization or accreditation agencies).

Conclusions

This study aimed to (1) measure departmental climate around teaching at the individual and department level using a newly developed instrument (DCaT) that has undergone rigorous validity and reliability studies and (2) explore the relationship between these measures of departmental climate and uptake of evidence-based instructional practices within the department. Analyses of surveys collected from 166 faculty members representing twenty-two STEM departments at research-intensive institutions show that (1) departmental climate around teaching is challenging to measure, and clear operationalization of what is being measured is necessary, and (2) measure of departmental climate around teaching at the

individual level (i.e., psychological collective climate) was not related to uptake of learner-centered instructional practices. This study is the first attempt at measuring departmental climate around teaching as defined in the organizational literature and testing a link between climate and instructional practices. Our findings suggest that departmental collective climate around teaching may be difficult to measure because most elements that define a climate (e.g., policies, practices, expectations) are lacking when it comes to teaching. Absence of these elements may contribute to the highly autonomous and independent approach to teaching that is seen in higher education and thus the lack of instructional innovation at scale.

Abbreviations

CFA: Confirmatory factor analysis; CFI: Comparative fit index; CSC NFW: Cottrell Scholar Collaborative New Faculty Workshop; DBER: Discipline-Based Education Research; DCaT: Departmental Climate around Teaching; ICC: Intra-class correlation coefficients; MIST: Measurement Instrument for Scientific Teaching; MLR: Robust Maximum Likelihood; MMC: Mixture model clustering; POGIL: Process-Oriented Guided Inquiry Learning; RMSEA: Root mean square error of approximation; SRMR: Standardized root mean square residual; STEM: Science, technology, engineering, and mathematics; WIDE R: Widening Implementation and Demonstration of Evidence Based Reforms; WLSMV: Weighted least square mean and variance adjusted

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40594-021-00303-1.

Additional file 1 : **Table S1.** DCaT Version 1. **Table S2.** DCaT Version 3. **Table S3.** DCaT Version 4 (final version). **Table S4.** Demographic questions. **Table S5.** Abbreviated Measurement Instrument for Scientific Teaching (MIST). **Table S6.** Constructs measured in prior studies on organizational climate. **Table S7.** Raw response rates by department. **Table S8.** Cleaned response rates for departments with highest raw response rates. **Table S9.** Estimated model proportions and estimated means for the VEE model (*n*=166). **Table S10.** Cognitive interview protocol. **Figure S1.** Academic rank comparison of whole department versus study participants for the three departments with the highest raw response rates. **Figure S2.** Tenure status versus type of psychological collective climate (*N*=166). **Figure S3.** Academic rank versus type of psychological collective climate (*N*=166).

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declaration

Competing interests

The authors declare that they have no competing interests.

References

Ab Hamid, M. R., Mustafa, Z., Idris, F., Abdullah, M., & Suradi, N. M. (2011). Measuring value–based productivity: A confirmatory factor analytic (CFA) approach. *International Journal of Business and Social Science, 2*(6), 85–93.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. American Educational Research Association.

Anderson, W. A., Banerjee, U., Drennan, C. L., Elgin, S. C. R., Epstein, I. R., Handelsman, J., … Olivera, B. (2011). Changing the culture of science education at research universities. *Science, 331*(6014), 152–153. https://doi.org/10.1126/science.1198280.

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. https://doi.org/10.1021/ed3002013.

Austin, A. E. (2011). *Promoting evidence based change in undergraduate science education*. National Academies.

Bathgate, M. E., Aragon, O. R., Cavanagh, A. J., Frederick, J., & Graham, M. J. (2019). Supports: A key factor in faculty implementation of evidence-based teaching. *CBE Life Sciences Education, 18*(2), ar22. https://doi.org/10.1187/cbe.17-12-0272.

Borda, E., Schumacher, E., Hanley, D., Geary, E., Warren, S., Ipsen, C., & Stredicke, L. (2020). Initial implementation of active learning strategies in large, lecture STEM courses: lessons learned from a multi-institutional, interdisciplinary STEM faculty development program. *International Journal of STEM Education, 7*(1), 4. https://doi.org/10.1186/s40594-020-0203-2.

Bouckenooghe, D., Devos, G., & Van den Broeck, H. (2009). Organizational change questionnaire–Climate of change, processes, and readiness: Development of a new instrument. *The Journal of Psychology, 143*(6), 559–599. https://doi.org/10.1080/00223980903218216.

Bouwma-Gearhart, J. (2012). Research university STEM faculty members' motivation to engage in teaching professional development: Building the choir through an appeal to extrinsic motivation and ego. *Journal of Science Education and Technology, 21*(5), 558–570. https://doi.org/10.1007/s10956-011-9346-8.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sociological Methods and Research, 21*(2), 230–258. https://doi.org/10.1177/0049124192021002005.

Brownell, S. E., & Tanner, K. D. (2012). Barriers to faculty pedagogical change: Lack of training, time, incentives, and… tensions with professional identity? *CBE Life Sciences Education, 11*(4), 339–346. https://doi.org/10.1187/cbe.12-09-0163.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83*(2), 234–246. https://doi.org/10.1037/0021-9010.83.2.234.

Childs, P. E. (2009). Improving chemical education: Turning research into effective practice. *Chemistry Education Research and Practice, 10*(3), 189–203. https://doi.org/10.1039/B914496J.

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology, 55*(4), 584–594. https://doi.org/10.1037/0022-006X.55.4.584.

Corbo, J. C., Reinholz, D. L., Dancy, M. H., Deetz, S., & Finkelstein, N. (2016). Framework for transforming departmental culture to support educational innovation. *Physical Review Physics Education Research, 12*(1), 010113. https://doi.org/10.1103/PhysRevPhysEducRes.12.010113.

Cottrell Scholars Collaborative. (2017). CSC New Faculty Workshop. Retrieved 02/04/2021 from http://chem.wayne.edu/feiggroup/CSCNFW/

Debad, S. J. (2020). *Recognizing and evaluating science teaching in higher education: Proceedings of a workshop*. National Academies Press. https://doi.org/10.17226/25685.

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), 31. https://doi.org/10.1186/s40594-018-0128-1.

Durham, M. F., Knight, J. K., & Couch, B. A. (2017, Winter). 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. https://doi.org/10.1187/cbe.17-02-0033.

Eagan, K., Stolzenberg, E. B., Lozano, J. B., Aragon, M. C., Suchard, M. R., & Hurtado, S. (2014). *Undergraduate teaching faculty: The 2013–2014 HERI faculty survey*. Higher Education Research Institute, UCLA.

Ehrhart, M. G., & Schneider, B. (2016). Organizational climate and culture. In O. Braddick (Ed.), *Oxford research encyclopedia of psychology*, (pp. 1–29). Oxford University Press. https://doi.org/10.1093/acrefore/9780190236557.013.3.

Elrod, S., & Kezar, A. (2016). *Increasing student success in STEM: A guide to systemic institutional change*. AAC&U.

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, 41(8), 578–588. https://doi.org/10.1093/comjnl/41.8.578.

Froyd, J. (2011). Propagation and realization of educational innovations in systems of undergraduate STEM education. In *Characterizing the Impact of Diffusion of Engineering Education Innovations Forum*. National Academy of Engineering.

Furnham, A., & Goodstein, L. D. (1997). *The organizational climate questionnaire (OCQ)*, (pp. 163–179). Pfeiffer Jossey-Bass, The Annuel.

Gess-Newsome, J., Southerland, S. A., Johnston, A., & Woodbury, S. (2003). Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching. *American Educational Research Journal*, 40(3), 731–767. https://doi.org/10.3102/00028312040003731.

Hage, J., & Aiken, M. (1967). Relationship of centralization to other structural properties. *Administrative Science Quarterly*, 12(1), 72–92. https://doi.org/10.2307/2391213.

Halpin, A. W., & Croft, D. B. (1963). *The Organizational Climate of Schools*. Chicago: Midwest Administration. New York: The Macmillan Company, Collier McMillan Ltd.

Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics-Physics Education Research*, 3(2), 020102. https://doi.org/10.1103/PhysRevSTPER.3.020102.

Henderson, C., & Dancy, M. H. (2011). Increasing the impact and diffusion of STEM education innovations. In *White paper commissioned for White paper commissioned for the Characterizing the Impact and Diffusion of Engineering Education Innovations Forum*.

Ho, G. W. (2017). Examining perceptions and attitudes: A review of Likert-type scales versus Q-methodology. *Western Journal of Nursing research*, 39(5), 674–689. https://doi.org/10.1177/0193945916661302.

Hora, M. T. (2012). Organizational factors and instructional decision-making: A cognitive perspective. *The Review of Higher Education*, 35(2), 207–235. https://doi.org/10.1353/rhe.2012.0001.

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. https://doi.org/10.1080/10705519909540118.

Hurtado, S., Eagan, K., Pryor, J. H., Whang, H., & Tran, S. (2012). *Undergraduate teaching faculty: The 2010–2011 HERI faculty survey*. Higher Education Research Institute: University of California, Los Angeles.

Kezar, A. (2016). Bottom-up/top-down leadership: Contradiction or hidden phenomenon. *The Journal of Higher Education*, 83(5), 725–760. https://doi.org/10.1080/00221546.2012.11777264.

Knorek, J. K. (2012). Faculty teaching climate: Scale construction and initial validation. In unpublished PhD dissertation, Urbana-Champaign: University of Illinois.

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. https://doi.org/10.1021/acs.jchemed.8b00220.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

Landau, S., & Chis Ster, I. (2010). Cluster analysis: Overview. In *International Encyclopedia of Education*, (pp. 72–83). https://doi.org/10.1016/b978-0-08-044894-7.01315-4.

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), 25. https://doi.org/10.1186/s40594-017-0092-1.

Lee, J. J. (2007). The shaping of the departmental culture: Measuring the relative influences of the institution and discipline. *Journal of Higher Education Policy and Management*, 29(1), 41–55. https://doi.org/10.1080/13600800601175771.

Lund, T. J., & Stains, M. (2015). The importance of context: An exploration of factors influencing the adoption of student-centered teaching among chemistry, biology, and physics faculty. *International Journal of STEM Education*, 2(1), 13. https://doi.org/10.1186/s40594-015-0026-8.

MacCallum, R. C., Widaman, K. F., & Sehee Hong, S. Z. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84.

Mamiseishvili, K., & Lee, D. (2018). International faculty perceptions of departmental climate and workplace satisfaction. *Innovative Higher Education*, 43(5), 323–338. https://doi.org/10.1007/s10755-018-9432-4.

National Science Foundation. (2013). Widening Implementation & Demonstration of Evidence-Based Reforms (WIDER). Retrieved 02/04/2021 from http://www.nsf.gov/pubs/2013/nsf13552/nsf13552.htm

National Science Foundation. (2020). Transforming higher education - Multidimensional evaluation of teaching. Retrieved 02/04/2021 from https://teval.net/index.html

Ngai, C., Pilgrim, M. E., Reinholz, D. L., Corbo, J. C., & Quan, G. M. (2020). Developing the DELTA: Capturing cultural changes in undergraduate departments. *CBE Life Sciences Education*, 19(2), ar15. https://doi.org/10.1187/cbe.19-09-0180.

O'Neill, T. A. (2017). An overview of interrater agreement on likert scales for researchers and practitioners. *Frontiers in Psychology*, 8, 777. https://doi.org/10.3389/fpsyg.2017.00777.

Ostroff, C. (1993). The effects of climate and personal influences on individual behavior and attitudes in organizations. *Organizational Behavior and Human Decision Processes*, 56(1), 56–90. https://doi.org/10.1006/obhd.1993.1045.

Patterson, M. G., West, M. A., Shackleton, V. J., Dawson, J. F., Lawthom, R., Maitlis, S., … Wallace, A. M. (2005). Validating the organizational climate measure: Links to managerial practices, productivity and innovation. *Journal of Organizational Behavior*, 26(4), 379–408. https://doi.org/10.1002/job.312.

Peterson, C. H., Peterson, N. A., & Powell, K. G. (2017). Cognitive interviewing for item development: Validity evidence based on content and response processes. *Measurement and Evaluation in Counseling and Development*, 50(4), 217–223. https://doi.org/10.1080/07481756.2017.1339564.

POGIL Team. (2019). POGIL. Retrieved 02/04/2021 from https://pogil.org/

Ramsden, P., Prosser, M., Trigwell, K., & Martin, E. (2007). University teachers' experiences of academic leadership and their approaches to teaching. *Learning and Instruction*, 17(2), 140–155. https://doi.org/10.1016/j.learninstruc.2007.01.004.

Reinholz, D. L., & Apkarian, N. (2018). Four frames for systemic change in STEM departments. *International Journal of STEM Education*, 5(1), 3. https://doi.org/10.1186/s40594-018-0103-x.

Reinholz, D. L., Corbo, J. C., Dancy, M., & Finkelstein, N. (2017). Departmental action teams: Supporting faculty learning through departmental change. *Learning Communities Journal*, 9, 5–32.

Robinson, O. C. (2013). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology*, 11(1), 25–41. https://doi.org/10.1080/14780887.2013.801543.

Schneider, B., Ehrhart, M. G., & Macey, W. H. (2013). Organizational climate and culture. *Annual review of Psychology*, 64(1), 361–388. https://doi.org/10.1146/annurev-psych-113011-143809.

Shadle, S. E., Marker, A., & Earl, B. (2017). Faculty drivers and barriers: Laying the groundwork for undergraduate STEM education reform in academic departments. *International Journal of STEM Education*, 4(1), 8. https://doi.org/10.1186/s40594-017-0062-7.

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance

structure models. *Journal of Business Research*, *58*(7), 935–943. https://doi.org/10.1016/j.jbusres.2003.10.007.

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., … Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science*, *359*(6383), 1468–1470. https://doi.org/10.1126/science.aap8892.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *J Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18.

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), 35. https://doi.org/10.1186/s40594-019-0185-0.

Subedi, B. P. (2016). Using Likert type data in social science research: Confusion, issues and challenges. *International Journal of Contemporary Applied Sciences*, *3*(2), 36–49.

The Coalition for Reform of Undergraduate STEM Education (2014). In C. L. Fry (Ed.), *Achieving systemic change: A sourcebook for advancing and funding undergraduate STEM education*. Association of American Colleges and Universities.

Thomas, K. W., & Tymon Jr., W. G. (2009). Work engagement profile. In *Cpp, Mountain View, Ca*.

Walczyk, J. J., Ramsey, L. L., & Zha, P. (2007). Obstacles to instructional innovation according to college science and mathematics faculty. *Journal of Research in Science Teaching*, *44*(1), 85–106. https://doi.org/10.1002/tea.20119.

Walter, E., Beach, A., Henderson, C., & Williams, C. (2014). Describing instructional practice and climate: Two new instruments. In *Transforming Institutions: 21st Century Undergraduate STEM Education Conference*, (vol. 24).

Wieman, C., Deslauriers, L., & Gilley, B. (2013). Use of research-based instructional strategies: How to avoid faculty quitting. *Physical Review Special Topics-Physics Education Research*, *9*(2), 023102. https://doi.org/10.1103/PhysRevSTPER.9.023102.

Wieman, C., Perkins, K., & Gilbert, S. (2010). Transforming science education at large research universities: A case study in progressxs. *Change: The Magazine of Higher Learning*, *42*(2), 6–14. https://doi.org/10.1080/00091380903563035.

Publisher's Note