RESEARCH   **Open Access**

Characterization of physics and astronomy assistant professors' reflections on their teaching: can they promote engagement in instructional change?

Haleigh Machost[1] , Emily A. Kable[1] , Jherian K. Mitchell-Jones[1] , Brandon J. Yik[1] and Marilyne Stains[1*]

Abstract

The development of reflective practitioners is one of four dominant change strategies in the Science, Technology, Engineering, and Mathematics (STEM) higher education literature. However, little research concerns the characterization of faculty's reflections. Before professional development programs can effectively incorporate reflective writings as a tool for pedagogical improvement, it is necessary to first understand the current state of faculty's reflections. To accomplish this goal, 98 physics and astronomy instructors were recruited from a teaching-focused professional development workshop and were asked to write a reflection on a self-identified challenging teaching experience.

A combination of *a priori* coding to analyze the content and depth of the reflections, as well as in vivo coding to better capture instructors' thinking were utilized. The majority of instructors wrote low-level reflections, wherein connections were not made between an instructors' actions and the observed outcomes or the described experience was not centered on students' outcomes or educational research literature. Approximately half of the instructors contemplated their own growth and the relationships with their students. However, only a small minority of instructors considered larger societal, cultural, or ethical factors. Plans created by instructors to address future, similar situations heavily relied on the instructors themselves, regardless of the depth of their reflections, and few planned to seek out knowledge from other resources such as peers or the education literature.

This study indicates that instructors may not engage in the types of reflection that are considered to promote meaningful instructional change. Trends in the instructors' plans show that ongoing support is necessary for them to effectively reflect and grow as practitioners. Overall, this work provides valuable insight into the poorly understood nature of faculty's reflections and showcases the need for more research to fully characterize reflections across STEM disciplines and to better inform professional development.

Keywords Reflection, Reflective practice, STEM instructors, Physics instructors, Astronomy instructors, Pedagogical training, Instructional reform

*Correspondence:
Marilyne Stains
mstains@virginia.edu
[1]Department of Chemistry, University of Virginia, Charlottesville, VA, USA

Springer Open

Introduction

Extensive evidence of the inequitable and poor learning outcomes experienced by students enrolled in science, technology, engineering and mathematics (STEM) courses (e.g., Hatfield et al., 2022; Koester et al., 2016; Matz et al., 2017) have reinforced calls for enhancing these learning environments. The need for such reforms has been recognized for decades by government bodies (Olson & Riordan, 2012), higher education organizations (Boyer Commission on Educating Undergraduates in the Research University, 1998; Miller & Fairweather, 2015), and STEM faculty themselves (Bradforth et al., 2015). Education researchers have responded by empirically investigating how students learn in STEM (e.g., Pond & Chini, 2017; Wu & Rau, 2019), examining the cognitive and affective challenges students experience in STEM courses (e.g., Marshman et al., 2018; Rice et al., 2013; Sorby et al., 2018), and leveraging findings from these studies to develop innovative instructional practices and test their efficacy on student learning (e.g., Chasteen et al., 2016; Henderson et al., 2011; Madsen et al., 2017; Mooring et al., 2016). As these evidence-based instructional practices emerged, different communities have strived to encourage STEM instructors to implement these practices in their courses. A review of studies on strategies to promote instructional change demonstrates the complexity of this endeavor (Henderson et al., 2011), and recent studies suggest that the uptake of these practices has been slow across STEM fields (Beane et al., 2019; e.g., Stains et al., 2018; Yik et al., 2022). The Henderson et al. (2011) review concluded that one important first step to change instructional practices is for instructors to understand their practices, beliefs, and values around teaching and to help them problematize their teaching. While this alone is not sufficient and long-term support and cultural change around teaching at the department and institution levels are also required, this step is essential as the dissatisfaction experienced once a problem is identified can be a powerful initiator for change (Andrews & Lemons, 2015).

Engaging STEM instructors in reflective teaching practices is a promising strategy to help them problematize their teaching. Indeed, reflections provide opportunities for instructors to critically analyze their teaching practices and learn from these analyses to enhance instructional effectiveness and, ultimately, students' experiences (McAlpine & Weston, 2002). The positive impacts of instructors' reflections have been reported extensively, especially in the K-12 literature (Ansarin et al., 2015; Belvis et al., 2012; Fox et al., 2011; Markkanen et al., 2020; Tajeddin & Aghababazadeh, 2018). In higher education, many of the calls for reforms on the evaluation of teaching and evaluation of teaching frameworks have also recognized the importance of reflections (Accelerating Systemic Change Network, 2023; Bradforth et al., 2015; Dennin et al., 2017; Simonson et al., 2022; The University of Kansas Center for Teaching Excellence, 2024; Weaver et al., 2020). For example, practicing reflective teaching is one of the four criteria described in the Framework for Assessing Teaching Effectiveness (FATE; Simonson et al., 2022) and an essential component of the Benchmarks for Teaching Effectiveness developed by the Center for Teaching Excellence at the University of Kansas (2024).

These teaching evaluation frameworks and guidelines are built on the premise that engaging in reflections will lead instructors to engage in instructional growth and the adoption of learner-centered practices. However, the literature on reflective practice has demonstrated that reflections can range in quality and therefore may not lead to expected outcomes (Dyment & O'connell, 2010; O'Connell & Dyment, 2011; Ryan, 2013; Spalding & Wilson, 2002). The teaching evaluation frameworks describe reflections in broad terms and provide limited scaffolding. For example, FATE describes an exemplary reflection as one that "demonstrates a high level of self-reflection around teaching broadly, objectively describing their strengths and weaknesses, consistent with evidence of teaching practices" (Simonson et al., 2022, p. 170). Similarly, the Benchmarks for Teaching Effectiveness describes someone with an expert level of reflection as an individual who "regularly adjusts teaching based on reflection on student learning, within or across semesters and examines student performance following adjustments" (The University of Kansas Center for Teaching Excellence, 2024). The literature on reflective practice has demonstrated that certain scaffoldings and methods are more effective at prompting high level reflections (i.e., reflections in which the instructor considers their roles, beliefs system and knowledge about teaching, and the place these play in the education of their students) and, therefore, at problematizing teaching. Unfortunately, few studies have explored the nature and quality of STEM instructors' reflections, whether as part of the teaching evaluation frameworks previously discussed or when instructors are provided with a specific, empirically-derived scaffold. It is necessary to first determine whether instructors are functioning as reflective practitioners on the level required to generate instructional change in order to design effective trainings and interventions involving reflective practice. Consequently, the goal of this study is to expand our understanding of the nature of STEM instructors' reflections by analyzing responses from physics and astronomy assistant professors to a specifically-designed reflective scaffold. The following research questions drive this study:

1. What is the nature of a difficult or challenging teaching experience (i.e., critical incident) new

postsecondary physics and astronomy instructors choose to reflect on?

2. What is the content of new postsecondary physics and astronomy instructors' reflective writings when prompted to consider a critical incident?
3. What depth of reflection do new postsecondary physics and astronomy instructors spontaneously reach?
4. What types of plans are new postsecondary physics and astronomy instructors proposing to address their critical incident?
5. To what extent are the nature of the critical incident, content of reflections, and plans outlined associated to the depth of these new postsecondary physics and astronomy instructors' reflection?

Reflective practice

Reflective practice has a history grounded in philosophy and the concept of reflective thinking, particularly in the work of Dewey (1933). The transition of reflective thinking to reflective practice- wherein the process of reflection is formalized and often recorded in some manner- lies in the realm of professional training, a shift which was catalyzed by the combined works of Schön (1983, 1987, 1991). Subsequently, Schön's concept of reflective practice has become extremely influential in the training of educators and healthcare professionals (Munby & Russell, 1989). Reflective practice is a process by which one considers past, present, or hypothetical experiences in light of personal belief system, assumptions, and knowledge base related to these experiences in order to gain insight concerning the factors at play as well as to plan for future, similar situations (Machost & Stains, 2023).

Reflective practice can be implemented through a variety of written, recorded, and oral methods (Machost & Stains, 2023). No matter the modality, the effectiveness of reflective practice stems from enabling instructors to deeply contemplate both their experiences and the knowledge they gained through these experiences (Machost & Stains, 2023; Osterman & Kottkamp, 2004). Indeed, by practicing continual and cyclical reflective practice, instructors can become more aware of their current pedagogical content knowledge and how they continually develop knowledge (Loughran, 2002). For this reason, reflective practice has been adopted as an important component of the professional development of educators (Marshall, 2019; McAlpine et al., 2004).

Reflection promotes greater effectiveness through encouraging planning for future experiences (Bain et al., 2002; Mohamed et al., 2022; Zahid & Khanam, 2019), focusing on one's strengths (Brookfield, 2017; Mohamed et al., 2022), and considering weaknesses and potential areas of improvement (Bain et al., 2002; Huda & Teh, 2018; Mohamed et al., 2022). In this way, reflection can problematize one's action and inspire the adoption of new approaches. Indeed, reflective practice is proposed to act as a "gyroscope" when navigating various external influences on the classroom, such as new departmental initiatives (Brookfield, 2017). Furthermore, it has been posited that "without routinely engaging in reflective practice, it is unlikely that practitioners in higher education will comprehend the effects of their inspirations, motivations, expectations and experiences upon their practice" (Lubbe & Botha, 2020, p. 290). For instance, through thoughtful reflection, instructors may realize how their own beliefs about the difficulty of a subject affect their explanations in class, or how their feelings of self-doubt affect their actions during office hours. Essentially, reflective practice acts as a magnifying glass, where instructors are able to analyze their actions and thoughts in relation to their experiences.

Analytical frameworks for reflections

Different frameworks have been presented in the literature to describe the nature and quality of reflections. Some frameworks focus on the variety of reflection types presented in one whole reflection (i.e., content), while others aim to evaluate hierarchically the depth of the reflection as a whole. The most popular frameworks leveraged in the literature that address these two aspects are presented below.

Content of reflections One predominant method of analyzing reflections is based on the content discussed within the reflection itself. This method originated with the work of Valli (1997). Within this model, there are five distinct types of reflection (Table 1): reflection-in and on-action, deliberative, technical, personalistic, and critical reflections. *Reflections-in and on-action* were derived from the work of Schön (1983) and relate to when the instructor is engaging in reflection, either while teaching (in-action) or after the act of teaching (on-action). *Deliberative reflections* are concerned with weighing different perspectives, opposing research findings, or varying personal viewpoints to determine the best course of action. *Technical reflections* are specifically concerned with following the guidelines put forth by a professional organization outside of the instructor; additionally, these guidelines must be based on pedagogical research to be considered technical-type reflection. *Personalistic reflections* involve "an educator's personal growth as well as the individual relationships they have with their students" (Machost & Stains, 2023, p. 5). Finally, *critical reflections* center on an instructors' own values, assertions, and assumptions about topics such as gender, accessibility accommodations, and cultural differences. Notably, the

Table 1 Content of reflections based on Valli (1997)

Content type	Definition
In- and On-Action	Content focused on an instructor's own past experiences, either retrospectively or in the moment.
Deliberative	Content centered on debating viewpoints, perspectives, or research, which are in opposition to each other. This content is associated with an instructor deciding on which pedagogical practices to change, alter, or implement.
Technical	Content centered on an instructors' pedagogical practices and the ways in which they control the classroom or teach their students. These considerations are performed in relation to guidelines created by an entity outside of the instructor, and the guidelines must be based on education literature.
Personalistic	Content centered on the relationships present in a learning environment. This includes the relationship between an instructor with their students, an instructor with their peers, and an instructor with themselves.
Critical	Content centered on how societal and cultural phenomena (such as gender, accommodations, and cultural differences) affect the learning environment.

Table 2 Depth of reflection based on Larrivee's (2008a) model

Level	Definition
Pre-reflection	Instructors base their teaching practices on pre-conceived notions, and do not comment about pedagogical goals they attempt to accomplish. There is a lack of connection between an instructor's actions and the observed outcomes.
Surface reflection	Instructors are concerned about achieving a specific goal, such as a specific passing rate for their class. However, these goals are only approached through conforming to departmental norms or their own anecdotal evidence. Thus, they are grounded in personal assumptions and influenced by unexamined beliefs and unconscious biases.
Pedagogical reflection	Instructors are willing to challenge the status quo and alter their pedagogical practices in light of evidence in observed student outcomes, relevant education literature, and alternative viewpoints. In this way, instructors also consider their own pedagogical belief system and its relationship to their practice.
Critical reflection	Instructors consider how societal and cultural phenomena affect the learning environment. In doing so, instructors evaluate their own views, assertions, and assumptions about teaching, with attention paid to how such beliefs impact their students holistically.

different types of reflection can occur simultaneously within the same piece of reflective writing.

Depth of reflections Reflective writings have been evaluated for depth through several different categorizations (Day, 1993; Farrell, 2003; Handal & Lauvas, 1987; Jay & Johnson, 2002; Larrivee, 2008a; van Manen, 1977; Zeichner & Liston, 1987). Larrivee (2008a) conducted an extensive review of this work in order to develop a four-level hierarchical model that represents the commonalities across these different categorizations (Table 2). Larrivee's model begins with *pre-reflection* where there is an absence of reflection. At the next level, we have *surface-level reflection* where an instructor is concerned about achieving a specific goal and also acknowledges a link between their actions and the observed outcomes; however, the desired outcomes are only approached through considering pedagogical norms, their own anecdotal experiences, or other practices established within the status-quo (Campoy, 2010; Larrivee, 2008a). In a *pedagogical-level reflection*, an instructor reflects on their educational goals and theories in light of observed outcomes in student comprehension, recent education research and literature, and alternative viewpoints (Larrivee, 2008a). Finally, *critical-level reflections* consider the ethical, moral, and political ramifications of what is being taught in an educational environment; furthermore, educators are evaluating "their own views, assertions, and assumptions about teaching, with attention paid to how such beliefs impact students" (Larrivee, 2005, 2008a; Authors, 2023, p. 4). The clear connection between critical-type reflection (re: content; Valli, 1997) and critical-level reflection (re: depth; Larrivee, 2008) should be noted. However, unlike content-based analyses of reflection, depth-based analyses are mutually exclusive. A piece of reflective writing is judged holistically and can only have one associated depth. Thus, while critical-type content is required for the critical-level to be reached, the presence of critical-type content does not automatically indicate a critical-level reflection. Additionally, a piece of reflective writing is associated with a depth, and the individual doing the reflecting is not bound to a particular level of depth; i.e., multiple reflections from an individual may have different associated contents and depths. It is important to note that instances of both pedagogical and critical reflections are considered by the authors to be high-level reflections.

Methods
This study, including participant recruitment, was approved by the Institutional Review Board for the Social and Behavioral Sciences at the University of Virginia (Protocol #: 5248).

Reflection scaffold
When conducting a review of the literature on reflective practice (Machost & Stains, 2023), authors HM and MS created a scaffold for written reflection based on the works of Gibbs (1988), Larrivee (2000, 2008a, b), and Bain et al. (2002); this scaffold was additionally inspired by other reflection scaffolds developed by the University of Edinburgh that were also based on some of this literature (The University of Edinburgh, 2021). This scaffold

begins by prompting participants to self-identify a past challenging teaching situation, i.e. a critical incident. Participants are then asked to describe the facts of the situation before being prompted to describe their feelings and the potential feelings of others involved. Next, they evaluate the critical incident for cause-effect relationships and positive/negative aspects, and finally draw conclusions from the critical incident and plan for future, similar situations. For each step of the process, an example of an answer to the scaffolding question was provided. The full scaffold is available in Appendix A.

Participants

Participants were recruited from two iterations of a national workshop for new physics and astronomy instructors (Physics and Astronomy Faculty Teaching Institute, 2023). Participants represented instructors at a variety of degree granting institutions (i.e., AA/AS, BA/BS, MA/MS, PhD) across all regions of the continental United States.

The first cohort of participants completed a Qualtrics survey containing the previously described scaffold during the workshop held in July 2022. The second cohort of participants completed the survey as a pre-workshop activity in June 2023. This change was implemented as the workshop itself was redesigned to heavily focus on reflection; thus, the pre-workshop survey serves as a baseline for participants' engagement in reflective practice prior to receiving instruction on reflective practice.

Participants were included in the study if they met the following criteria: (1) the reflection submitted had to be about a time or situation when the participants were acting as an instructor; and (2) the description of the critical incident had to be clear and detailed so as to (i) be easily understood and (ii) not require interpretation by the research team. Of the 62 instructors who attended the July 2022 workshop, 52 submitted a reflection, and 46 met the inclusion criteria. Of the 106 instructors who attended the June 2023 workshop, 57 submitted a reflection, and 52 met the inclusion criteria. A total of 98 reflections were included for analysis.

Scaffold analysis

A combination of in vivo and *a priori* coding was used in the creation of the codebook. The codebook is comprised of four sub-codebooks containing codes generated to describe the following categories: topics discussed in the reflections, content of the reflections, level of the reflections, and plans created in the reflections. Two code categories, topics and plans, were created solely from in vivo coding. None of these in vivo codes were mutually exclusive within each code category or across the different code categories. Two code categories, content and level,

were created *a priori* from Valli's (1997) and Larrivee's (2008a) descriptions of content and depth, respectively.

Topic codes were used to capture the nature of the critical incident. In all, 18 topic codes were used by authors HM, EAK, JKMJ, and BJY during coding and assessment of inter-rater reliability; after inter-rater reliability analyses, these 18 topic codes were condensed into 9 parent-categories following analysis by authors HM and MS (Table 3).

The plan codes were used to capture the actions participants either had taken or plan to take to prepare themselves for future, similar situations. In all, 25 plan codes were utilized by authors HM, EAK, JKMJ, and BJY. Post inter-rater analyses, only the plan codes utilized in at least 5% of the written reflections were retained for further analysis. Authors HM and MS organized these remaining 15 plan codes into three categories based on the intent behind each individual plan. For a full summary of the plan codes and their utilization frequency, see Appendix C, Table S2.

The portion of the codebook used to describe the content of reflections, as depicted by Valli (1997), was created using a mixture of *a priori* and in vivo coding. As other analyses of reflection have done (Minott, 2008), Valli's five categories were utilized in *a priori* coding. However, these five categories were each expanded upon with subcodes derived from in vivo coding to give a better understanding of the content described (see Results and Discussion). As with the topics and plans codes, the content-based codes were not mutually exclusive either across or within code categories.

Finally, the portion of the codebook depicting depth of reflection used *a priori* coding taken from Larrivee's (2008a) description of the different levels of reflection. Larrivee's four-level categorization has previously been used in the analysis of reflections (Ansarin et al., 2015; Campoy, 2010), and a modified version of Larrivee's categorization has also been used (Winchester & Winchester, 2011). However, other analyses use a different depth-model of reflection (Betrabet Gulwadi, 2009; Dyment & O'Connell, 2010; Jensen & Joy, 2005; Lee & Abdul Rabu, 2022; O'Connell & Dyment, 2004; Plack et al., 2005; Richardson & Maltby, 1995; Sumsion & Fleet, 1996; Thorpe, 2004; Wong et al., 1995). The categorization used herein is described in the introduction and aligns with Campoy's (2010) and Larrivee's (2008a) works. As the depth of reflection is a holistic analysis, these codes were mutually exclusive. Importantly, for a reflection to be classified at the critical level, the higher-level concerns (e.g., equity, accessibility, representation, etc.) must have been considered consistently throughout the entirety of the reflection.

Table 3 Topic categories, topic codes, and definitions

Topic category	Topic code	Definition
Poor student(s) behavior	Student in-class disruption	The instructor describes a situation where students are being disruptive (e.g., talking during lecture, arguing before class, etc.)
	Student cheating	The instructor describes a situation where a student cheated on an exam/assignment
	Student procrastination	The instructor describes a situation where a student is procrastinating
External to professor or class	Equipment failure	The instructor describes a situation where the necessary equipment fails during class or lab
	COVID transition	The instructor describes a situation dealing with the COVID transition to online classes or coming back to in person instruction
Class management	Made assessment too difficult	The instructor comments on an assessment or assignment they designed that was too difficult for their students (either due to time constraints or just the complexity of the material)
	Poor class time management	The instructor describes a situation during which they moved too quickly through a class, had too much material expected to be covered in a class, etc.
	Recommendation letter	The instructor reflects on a situation that arose while writing a recommendation letter
Student(s) negative feedback	Student direct negative feedback	The instructor describes a situation where students complain to the instructor directly or via a feedback survey
	Student indirect negative feedback	The instructor describes a situation where students' complained about the instructor/course to others (e.g., colleagues) or via end-of-course evaluations
Critical topics	Sexually inappropriate behavior	The instructor describes a situation where sexual harassment or actions contributing to sexual harassment were taking place in the classroom
	Cultural differences	The instructor describes a situation during which cultural differences contributed to difficulties experienced by either the student or the instructor
	Gender	The instructor describes a situation where gender norms, roles, or expectations play a part in the learning environment. The role of gender may be explicit in the description or assumed by either the student or instructor
Students' weak academic profile	Students lack fundamentals	The instructor describes a situation where students have a weak understanding of fundamental concepts and skills
	Student poor performance	The instructor describes a situation where a student is not performing well academically in class/lab
Struggling student(s)		The instructor describes a situation where a student is not doing well holistically
Student-instructor specific interactions		The instructor describes a difficult student interaction, including combative interaction on the part of the student, correcting students' behaviors in class, or the instructor being abrasive
Instructor's incorrect answer or explanation		The instructor describes the situation where they gave an incorrect answer or explanation or were not able to give any answer or explanation

Trustworthiness

Steps were taken throughout this analysis to ensure credibility, transferability, and dependability.

Credibility As outlined by Shenton (2004), there are numerous avenues to demonstrate credibility. First, we approached the analysis by adopting "research methods well established" in the literature (Shenton, 2004, p. 64); *a priori* coding taken from the well-established works of Valli (1997); Larrivee (2008a) aided in ensuring that the analysis aligns with prior work when determining the content and depth of the reflections. Furthermore, we address the previous findings in the literature while discussing the findings from this study.

Throughout the analysis of the data, frequent debriefing sessions occurred within the entire research team. Finally, we aim to establish transparency of both the data and the data analysis through the information provided in Appendix B, Table S1.

Transferability We promoted transferability by providing a thick description of the context of the study, its participants, the data collection, and analysis processes.

Dependability The stability of our findings is primarily addressed via the two different samples, collected one year apart. Similar distributions of content and depth were seen at the two different time points, and the codebook developed after the first data collection readily applied to the second set of data. The initial codebook was created through an iterative code-recode strategy by author HM informed by whole-group discussions with the research team. Additionally, the final codebook demonstrates inter-rater reliability with percent agreements greater than 80% in all code categories; 16 of the 46 reflections from the initial data collection were fully cross coded between HM, JKMJ, and BJY to demonstrate reliability (Table 4). Due to the non-mutually exclusive nature of the codebooks, Cohen's kappa values were not calculated.

This cross-coding was performed through stepwise replication across five rounds. Furthermore, throughout the initial sense-making and the intensive inter-rater reliability analyses, a detailed audit trail was kept about the iterative modifications of the code books. Changes made

Table 4 Inter-rater reliability metrics. All codebooks were fully cross-coded by A, C, and D

Codebook	Number of interviews	Percent agreement
Topics	16/46	83%
Content	16/46	89%
Level	16/46	89%
Plans	13/46	87%

Table 5 Distributions of topics discussed in critical incidents

Topic	Proportion of reflections
Student(s) weak academic profile	39%
Student-instructor specific interactions	28%
Student(s) negative feedback	26%
Instructor's incorrect answer or explanation	17%
Poor student(s) behavior	14%
Class management	8%
Critical topics	7%
Struggling student(s)	6%
External to professor or class	6%

during the first three rounds of inter-rater reliability analyses include: altering the names of codes (e.g., changing 'student indirect complaints/evaluations' topic code to 'student indirect feedback'), adding onto the definitions of codes (e.g. definition of 're-explain course material' planning code was expanded to explicitly include utilizing a different method or approach), and verbal clarifications (e.g. that not all sections of the codebook needed to be utilized in each reflection). The final two rounds of inter-rater reliability analyses resulted in no further changes to the codebook. For a detailed analysis of the changes resulting from the iterative inter-rater reliability analyses, see Table S1.

Results and discussion

The findings discussed herein provide insight into the written reflections of physics and astronomy assistant professors who are untrained in reflective practice. The presentation of the results is aligned with the research questions.

Nature of critical incidents

Participants focused their critical incident on nine different topics (Table 3). The top three topics most discussed were *Student(s) weak academic profile*, *Student-instructor specific interactions*, and *Student(s) negative feedback* (Table 5). The following three excerpts provide examples for each of these three topics, respectively:

> *I was teaching a grad student class. Before the midterm, nearly the entire class was shaking their hand in agreement when I tried to gauge the clarity of my lectures. I did ask questions and encouraged different people to participate, but the first midterm performance was extremely poor and revealed a knowledge gap that I didn't expect to see.–Instructor 129.*

> *When I asked a question from a student to increase her engagement in class, she didn't answer. I helped her to get to the answer, but she didn't show any interest either. I provided the answer and asked her to make sure that she understood the process of getting to the answer, and she said," I will just say' yes'". It was clear her 'yes' was only to make me to leave her alone.– Instructor 114.*

> *I got very poor course evaluations and students made complaints to the department on grading. However, I asked students to talk to me at the very beginning of the semester if they have questions on their grades, and no one talk[ed] to me during the semester.–Instructor 138.*

Overall, the data indicate that most instructors' reflections were focused on negative events with students. Indeed, 79% of the critical incidents contained at least one topic code about negative experiences with students. At the time of the writing of this manuscript, we could not find studies that had explored the focus of teaching reflections written by higher education instructors in STEM and other disciplines. This study thus provides a first insight into what STEM instructors consider challenging situations within their teaching.

Content of reflections

We leveraged Valli's (1997) framework to analyze the content of the reflections, which includes five types (Table 1): in- and on-action, deliberative, technical, personalistic, and critical. Since the scaffold used to guide the written reflections requires the participants to reflect on a past teaching experience, the in- and on- action content type was not relevant to code.

Neither technical nor deliberative content were present in any of the participants' reflections. The lack of technical content aligns with a prior study investigating pre-service teachers enrolled in a course that required students to maintain a reflective journal throughout the term (Minott, 2008). In Minott's study, reflections were collected from 20 pre-service teachers where participants submitted five entries from their reflective journals for assessment. In these submissions, there were no instances of technical content, mirroring the findings from the present study. Importantly, Minott's participants had months to record reflections in a journal and chose which of their reflections to submit. Our study collected spontaneous reflections from participants without training in reflective practices. Thus, the lack of technical reflection in either participant pool may

indicate that technical reflection needs to be deliberately prompted. Unlike what is observed in our study, Minott (2008) noted instances of deliberative content in 10% of the study sample. This difference may be due to the scaffold used in our study (see Appendix A), which does not directly probe instructors to consider opposing perspectives or viewpoints.

Personalistic content was the most common content present in the reflections with 57% (*n*=56) of instructors addressing it. The prevalence of personalistic content aligns with Minott's (2008) prior study, as personalistic content was the second-most prevalent content type among Minott's participants, only surpassed by in- and on-action. We identified six subcodes that fit within personalistic content (Table 6). Our participants reflected mostly on themselves and their flaws or on negative perceptions that they thought others had about them. Few considered their students' holistic improvement or empathized with them, two key criteria for personalistic content (Authors, 2023; Minott, 2008; Valli, 1997).

Critical content was observed in significantly fewer reflections (12%, *n*=12) and felt into one of four subcodes: (1) Accommodations, (2) Gender, (3) Cultural differences, and (4) Grouping (Table 7). The presence of critical content in a minority of participants again aligns with Minott's (2008) findings, who noted critical content in only 3% of the reflections in their study. The most common critical content written about by our participants related to the need to accommodate students.

Depth of reflections

The depth of the reflections collected were analyzed using the four-level hierarchical categorization of reflections developed by Larrivee (Tables 2 and 2008a). Over 80% of the reflections written by our participants felt to the low-level of reflection, with 23 reflections classified at the pre-reflection level and 59 at the surface-level (Fig. 1). A hallmark of pre-reflection was a lack of connection between an instructor's actions and words and the observed outcome. This is exemplified with Instructor 108:

> *There was a girl in the class who did very well in almost all the homework. She never came to the office hour. But since she did well in homework, I thought she understood the materials well. But she didn't do well in mid-term. I discussed the midterm with her, and she told me that she had schedule conflict with the office hour. I then offered very flexible time to her, but she then never came. She continues to do okay on her homework until she did very poorly on the final... I feel confused about her performance on homework and exam. It seemed that she was cheating on her homework.–Instructor 108.*

Instructor 108 saw themselves as a bystander; they failed to see any reason for the conflicting performance of their student other than cheating. Additionally, there is no connection between a minimal action on the instructor's part and the student's continued mixed performance. This contrasts to surface-level reflection where instructors do make a connection between themselves and the outcomes; however, the plans to achieve different outcomes in surface-level reflections are based on anecdotal experiences or the status quo as Instructor 102 illustrates:

Table 6 Personalistic content subcodes, definitions, exemplary quotes, and distribution of subcodes within the reflections that contained personalistic content

Personalistic content subcode	Definition	Exemplary quote	Proportion of personalistic content reflections
Failure to facilitate learning	Instructor reflects on their inability to facilitate their students' learning	"I felt I failed the students on properly introducing them to a key concept in the course and felt like I was not a good teacher."–Instructor 257	54%
Perceived negative opinions of instructor by students	Instructor reflects on how they perceive their students to view them or the course	"I immediately felt a sense of dread and panic - thinking that my students would think I was a fraud."–Instructor 209	36%
Negative personal traits	Instructor reflects on their own negative personal traits (short temper, insecurities, etc.)	"I think it also reflected my own insecurities. I always had a bit of imposter syndrome, especially in grad school, so any sort of criticism of my teaching made me very defensive."–Instructor 134	16%
Failure as advocate	Instructor reflects on their inability to advocate for their students or their failure while advocating for them	"No one had prepared me for what I should do when a student starts having a breakdown/crisis in the middle of class. After the student left, I was mostly concerned that the student would be able to get help. I hope the student felt supported."–Instructor 249	12%
Peer interpretations or opinions of instructor	Instructor reflects on how they perceive their peers to view them or the course	"My colleague observing me definitely pitied me and tried to offer helpful suggestions." –Instructor 233	9%

Table 7 Critical content subcodes, definitions, exemplary quotes, and distribution of subcodes within the reflections that contained critical content

| Critical content subcode	Definition	Exemplary quotes	Proportion of critical content reflections
Accommodations	Accommodations for students with disabilities and/or difficult situations. May relate either to the implementation of the accommodations themselves or to how accommodations affect non-accommodated students.	"Some students, especially those with a family, might be too busy to do two homework assignments per week and to commit to outside class activities. That gave me a new understanding to accommodate everyone in class."–Instructor 204	58%
Gender	Gender adding a more complex layer to situations. Most often, it is the result of a female existing in a male-dominated field or class. The effects of this gender discrepancy can be expressed as either experienced by the student or be expressed as a concern by the professors navigating a situation	"A student neglected to write their pronouns in my get-to-know-you survey and asked me not to bring up their gender again. I was happy to respect that, but then they asked me to write them a letter of rec. I needed to write their pronouns in the letter, and I didn't know what to use."–Instructor 101	17%
Cultural differences	Cultural differences contributing to difficulties experienced by either the student or the instructor	"He was previously educated in another country, where the students were not able to ask questions (as that generally was viewed as meaning they weren't able to do things themselves). So I realized that when I told the class that I expected them to come talk to me about things they didn't understand, he still didn't think it was really an option." –Instructor 147	17%
Grouping	Grouping students together without reason (note: NOT due to race, gender, ethnicity, sexuality)	"I think things went poorly because I painted half of the class with a generalization. I realized afterwards that it would be better to address students' resistance to participation when they were in smaller groups, or perhaps individually."–Instructor 112	17%
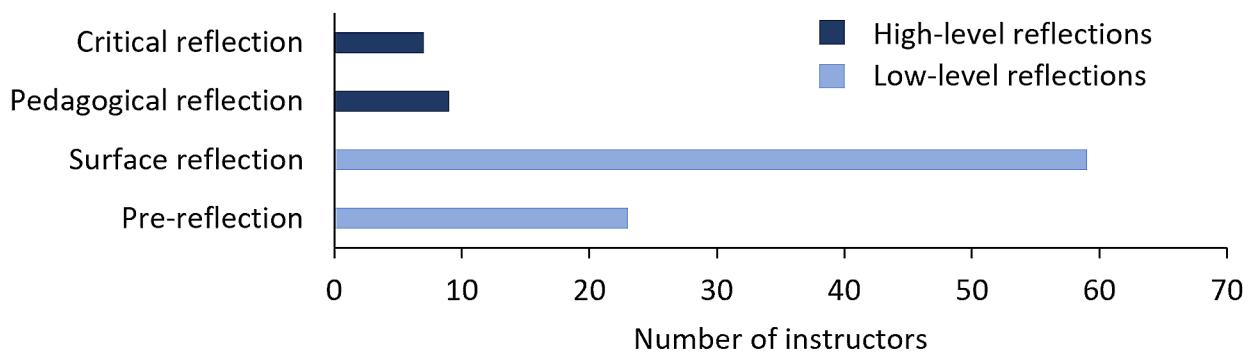
Fig. 1 Distribution of physics and astronomy instructors' reflections across the different levels of depth of reflection based on Larrivee's (2008a) model

"I learned that I need to be more prepared for my lectures, although this is an ongoing challenge for me. I do need to learn to handle my own mistakes with more grace. I'm OK with admitting that I'm wrong or don't know something, but I do that too much in my lectures."–Instructor 102.

A minority of participating instructors (16%) completed high-level reflection (Fig. 1): pedagogical-level reflection (*n*=9); critical-level reflection (*n*=7). As seen with Instructor 128, instructors who reached the pedagogical level focused on how they can improve their teaching based on observed outcomes in student comprehension, alternative viewpoints, and/or current educational research and literature:

"I learned that my style of sort of more casual research instruction… does not always help my student. I think I should learn more about teaching scientific programming to undergraduates, and what are some successful strategies or techniques I can impart to them. Hopefully in the future I'll be better prepared because I will have developed structured mini-lessons on best coding practices, and my stu-

dent and I will have done those together, and I will have also developed mechanisms for soliciting specific feedback from my students on what I can do to help them better learn."–Instructor 128.

Instructor 128 took from their experience that they need to change the status quo of how they taught coding to researchers. In doing so, they exhibit a high-level of reflection regarding their pedagogical practices. An added layer of complexity is present in those instructors who reach critical-level reflection as they examine the role that larger societal issues, trends, and differences play in learning environments. Instructor 147 details this relationship, as they had a student who had the potential to perform better in their course but did not do so because of cultural differences where the student was not comfortable asking questions. Furthermore, rather than problematizing the student, Instructor 147 acknowledges that it is their role as the instructor to make the classroom norms easily understood by students.

"I was working under the assumption that when I told students that not only could they ask questions and/or come to me for help, [that] they accepted it when I made the offer. This situation made me understand that some students (especially from certain backgrounds) had preconceived notions about what they should do as students, and that I needed to do more to encourage them."–Instructor 147.

These findings may seem in contrast to a prior study investigating Iranian English as a Foreign Language teachers which found the predominant depth of reflection among these instructors to be at the pedagogical level. Importantly, the researchers found a positive correlation between an instructor's years of teaching experience and the depth of their reflection (Ansarin et al., 2015). In their study, instructors had a broad range of teaching experience with an average of 8.39±4.59 years. In contrast, our study sample had less teaching experience; based on the demographic data that were collected from the 2023 cohort (no such data was collected from the 2022 cohort), the 2023 cohort had an average of 3.2±4.8 years of teaching experience. Therefore, our cohort is more similar to the group of instructors in the Ansarin et al. (2015) study who were classified in the low level of teaching experience. That group wrote a significantly larger proportion of pre-reflection and significantly less pedagogical and critical reflections. In light of these results, it may be that our sample provided fewer high-level reflections because they had not had enough teaching experiences.

Plans to address similar situations in the future

Instructors were asked to describe their plans for preparing themselves better when faced with a similar challenging situation in the future. Through in vivo coding, three majors plan codes emerged (Table 8): *Self-preserving*, *Self-reliant*, and *Seeking knowledge* outside of self. *Self-preserving* plans entail emotional regulation regarding either oneself or others (i.e., personal grace), or standard practices of instructors (i.e., pre-planning). *Self-reliant* plans go beyond the explicit duties of instructors and are based solely on an instructor's own experiences, speculations, and abilities to address the topics at hand (e.g., establishing clear expectations in the classroom, correcting mistakes made by oneself, meeting students where they are academically, discussing issues privately or in small groups). *Seeking knowledge* plans rely on an instructor going outside of their current knowledge or personal past experiences, and include soliciting student feedback, implementing successful strategies (either in the literature or as used by peers), communicating with peers, and participating in professional development. Instructors' plans relied mostly on the instructors themselves and their own knowledge and experiences (Table 8).

Only about a quarter thought to reach out and leverage other resources (e.g., peers, books, and peer-reviewed journal articles) to better equip themselves to handle future challenging situations. The nature of the plans presented in these reflections indicate that the engagement in the reflection is unlikely to lead to pedagogical growth among the participants.

Relationship between the nature of the critical incident and depth of reflections

We analyzed the relationship between the nature of the critical incident (i.e., topics; Table 3) and the depth of the reflection to explore whether certain situations are more prone to engage instructors in higher-level reflections. Table 9 displays the distribution of the topics explored in the critical incidents across the four levels of depth of reflection described by Larrivee (2008a; Table 2).

The topic of *Student(s) weak academic profile*, which was the most common topic discussed by our participants (Table 3) is equally represented across all levels of depth. Therefore, reflecting on students' academic difficulties can but does not necessarily lead to high-level reflections.

The topics that most distinctively separate reflections at the critical level from other levels were *Student-instructor specific interactions* and *Struggling student(s)*. The *Student-instructor specific interactions* were over twice as prevalent in the critical reflections than in the other levels of reflection. However, no notable qualitative differences were found between the descriptions of *Student-instructor specific interactions* at the critical level and

Table 8 Types of plan described in the reflections for managing future similarly challenging situations. Only plan subcodes present in at least 5% of the reflections were analyzed and are presented in this table. For a full list of planning subcodes and definitions, see Additional file 1: Appendix C, Table S2

Types of plan	Example subcode	Exemplary quote(s)	Proportion of reflections
Self-reliant	Communicate with students–establish clear expectations	"I learn[ed] from this situation that it is super important to set up classroom culture and be more attentive to class dynamics. I think in the future, I could integrate the activity of building community agreement at the beginning of the course and continue to revisit it throughout the semester to remind students of ways of working that they are expected to do and agreed to do"–Instructor 125	58%
	Re-explain course material	"I would step back and engage with some basic cross-product concepts that students know from mathematics, and then, once they get familiar, they move on to the idea they need to know."– Instructor 237	
Self-preserving	Pre-planning	"I've learned that preparation is important to avoid stressful and regretful situations (for both the instructor and the students)."– Instructor 232	45%
	Personal grace	"But I also realized I needed to give myself grace for not taking more frequent/detailed notes the first time I was teaching as it was ridiculously busy semester."– Instructor 236	
Seeking knowledge	Participate in professional development	"I have had to develop leadership skills to address interpersonal issues more directly, including attending workshops and taking courses on equity, inclusion, and social justice." – Instructor 121	27%
	Communicate with peers	"I am always looking for advice from teachers with experience running extra large classrooms."– Instructor 250	
No plan	Did not write a plan	"Students sometimes seem to assume (based on their predicted grades in the teaching evaluations) that I will curve more than I do. I worry this affects the amount of work they put into the class."– Instructor 216	9%
Other types of plan	Plans that were present in 5% or less of the reflections	"I don't know what to do under this situation."– Instructor 108 "I am unlikely to address questions that aren't strictly about content at my university ever again, which I think is a loss for both the students and for me."– Instructor 127	5%

Table 9 Distribution of topics discussed across the four levels of depth. Cell percentages represent the proportion of reflections at a specific level (i.e., depth) of reflection that included each topic category. Topic categories are not mutually exclusive; thus, the sum within a level is greater than 100%

Topic (from most to least reported)	Pre-reflection (n = 23)	Surface (n = 59)	Pedagogical (n = 9)	Critical (n = 7)
Student(s) weak academic profile	43%	36%	44%	43%
Student-instructor specific interactions	30%	25%	---	71%
Student(s) negative feedback	22%	30%	33%	---
Instructor's incorrect answer or explanation	9%	24%	11%	---
Poor student(s) behavior	26%	12%	11%	---
Class management	---	10%	22%	---
Critical topics	9%	7%	11%	---
Struggling student(s)	9%	2%	---	43%
External to professor or class	4%	5%	11%	14%

lower levels. Therefore, similarly to the *Student(s) weak academic profile* topic, the focus on student-instructor interactions does not seem to drive the depth of the reflection. The *Struggling student(s)* topic, which is when instructors are considering their students who appear to be struggling holistically rather than solely as students or academically, was only present in 7 of the 98 reflections,

but half of these reflection were at the critical level. The presence of this topic is in alignment with the definition of critical level by Larrivee (2008a). However, it is worth noticing that few of the lower-level reflections covered this topic as well. While these instructors had described students struggling holistically, they did not make it the focus of their reflections and were thus not classified in the higher-level of reflections. This points to a missed opportunity for instructors to engage in more transformative reflections but also indicates that instructors need to be guided towards unpacking more this type of topics.

Overall, the data in Table 9 do not provide a clear trend (except for *Struggling student(s)*) between the topic being discussed in the critical incident and the depth of the reflection. This finding indicates that it may not be necessary to coach faculty to think about particular types of situations in order for them to engage in high-level reflections. Other aspects, such as the content of the reflection, might play a bigger role and will be explored in the next section.

Relationship between the content and depth of reflections
While the connection between content and depth of reflection may appear to be intuitive, few studies simultaneously analyze reflections for both content and depth (e.g., Lee, 2005). This is an important gap in the literature as understanding the content that appears in high-level
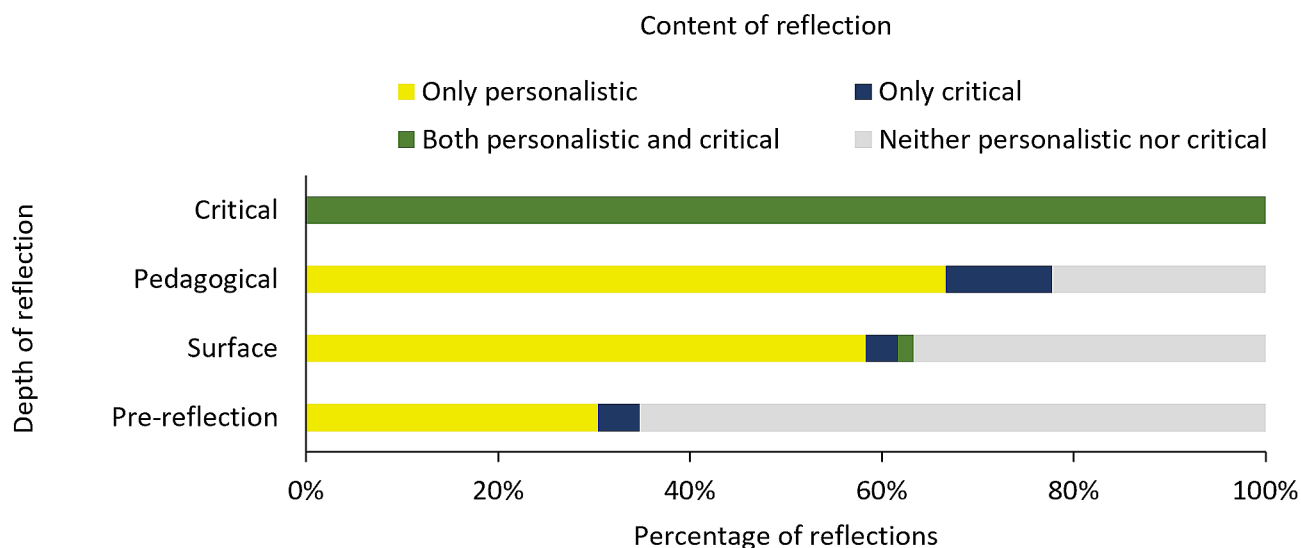
Fig. 2 Overlay of content and depth of instructors' reflective writings. Percentages are normalized for each level of reflection

Table 10 Distribution of plans described by instructors across the four levels. Cell percentages represent the proportion of reflections at a specific level of reflection (i.e., depth) that included each type of plan. As instructors could describe multiple types of plan within the same reflection, the sum within a level is greater than 100%

Type of plan (from most to least reported)	Pre-reflection ($n = 23$)	Surface ($n = 59$)	Pedagogical ($n = 9$)	Critical ($n = 7$)
Self-reliant	43%	56%	89%	86%
Self-preserving	17%	56%	44%	43%
Seeking knowledge	30%	24%	67%	14%
No plan	17%	8%	---	---
Other types of plan	13%	3%	---	---

reflections can aid in the development of reflective practitioners.

Figure 2 depicts the relationship between the content and depth of reflections. As the level of reflection increases so does the presence of personalistic content. At the pre-reflection level, most reflections contain neither personalistic nor critical content, while all critical reflections contain both personalistic and critical content. Interestingly, critical content is a distinctive feature of reflection at the critical level since it is mostly absent in the pre-reflection, surface, and pedagogical reflections. Therefore, it is essential to guide instructors towards exploring critical content (e.g., gender, accommodations, and cultural differences) when they engage in reflection. However, as the presence of critical content in the low-level reflections indicates, it might not be sufficient. Similar to our previous recommendations about guiding instructors to further unpack the topic of *Struggling student(s)*, instructors also need to be guided in exploring critical content for them to reach higher level reflections.

Relationship between the plans outlined and depth of reflections

Each type of plan (i.e., *Self-reliant*, *Self-preserving*, and *Seeking knowledge*) was observed across all depth levels (Table 10), but each level of reflection had a different combination of plans (Fig. 3 and Appendix C, Table S3).

Low-level reflections contained more diverse plans and were more likely to have a combination of plan types when compared to high-level reflections. However, low-level reflections were also the only reflections for which the *No plan* code was used, albeit at a small rate (Fig. 3). The most common type of plans in each of the low-level reflections was *Self-reliant (*Table 10*)*. Reflections in both the pre-reflection and surface reflection levels also had roughly a quarter of the plans focused on *Seeking knowledge*. What clearly differentiated the two low-levels of reflection was the proportion of *Self-preserving* plans, which was higher in the surface reflections when compared to the pre-reflections.

High-level reflections had limited types of plans and were dominated by *Self-reliant* plans (Table 10). Nearly half of the reflections at the high reflection levels also included *Self-preserving* plans. A key distinction between the pedagogical and critical levels was the much higher
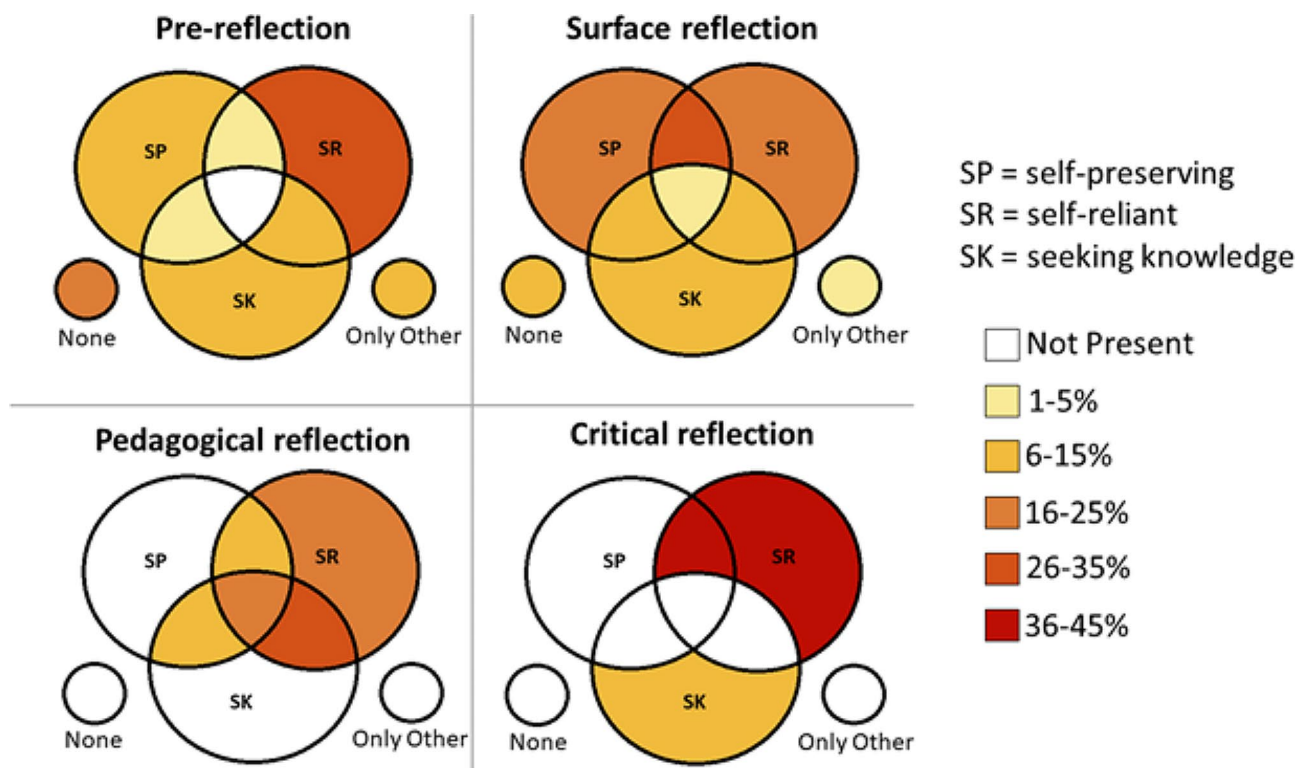
Fig. 3 Combinations of plans among the four levels of reflection

proportion of *Seeking knowledge* plan in the pedagogical reflections (67% versus 14%, respectively). Indeed, the critical level had the smallest proportion of reflections with *Seeking knowledge* plans (14%); this could be due to the difficult subject matters broached in the critical-level reflections which instructors may be hesitant to discuss with outside sources.

Overall, the data show that regardless of the level of reflection, instructors rely on themselves to prepare for the next time they face a similar critical incident. Therefore, instructors' engagement in these reflections are not likely to result in pedagogical growth. Our data indicates that we need to normalize seeking help from others when facing challenging teaching situations. A recent study that qualitatively explored the teaching social network of STEM faculty had probed help-seeking behaviors of STEM instructors when faced with issues with their teaching (Lane et al., 2022). They found that many of the 19 interviewees would only reach out to their discussion partner if they knew that this instructor had the expertise and experience that was directly related to the problem they were encountering. This current study and the Lane et al. study (2022) demonstrate the need to promote communications among instructors so that they can learn about the breadth of expertise of their peers, and thus have resources that they can feel comfortable reaching out to when facing a challenging situation.

Implications

Findings from this study lead to several implications regarding the promise of reflective practice in promoting pedagogical growth and the research agenda around reflective practice.

The required inclusion of reflections on teaching evaluations is likely not enough to promote pedagogical growth: instructors need to be trained on reflective practice

This study showed that instructors with limited teaching experience wrote low-level reflections. Low-level reflections mean that instructors are not considering their beliefs and values about teaching, nor educational literature when reflecting on a critical incident. Our data also show that instructors are primarily looking inward when elaborating plans to address future similar situations. As Henderson et al. (2011) remarked in their review of the literature on instructional change, it is essential for instructors to face their beliefs/values around teaching in order to better problematize their teaching. Moreover, their self-reliance is unlikely to lead these instructors towards learning new instructional approaches or ways of supporting their students. Consequently, the instructors in this study are unlikely to experience pedagogical growth as a result of their writing of these reflections.

As indicated in the introduction, reflections are becoming a center-piece of new teaching evaluations and are

seen as a mean to help instructors improve their teaching practices (Simonson et al., 2022; The University of Kansas Center for Teaching Excellence, 2024). Our data suggest that this requirement alone is insufficient to achieve this goal and that training instructors is necessary. This is also in-line with prior research on reflective practice (Belvis et al., 2012; Dinham et al., 2021; Zahid & Khanam, 2019). Our data points to the need to train instructors in recognizing and unpacking critical topics and in considering students more holistically. Trainings should also provide instructors with educational resources and trusted networks of pedagogically-trained colleagues that they can leverage to gain insight about their particular situation and identify strategy to mediate similar future situations.

A more extensive research agenda around STEM instructors' reflective practice is needed to design effective training

This study is one of the first studies to characterize the nature of STEM instructors' reflections on teaching. Consequently, more studies ought to be conducted to characterize the generalizability of these results across STEM fields (we only have physics and astronomy instructors in this study) as well as a range of teaching experiences and contexts (e.g., type of course, class size, type of institution). Extending this research agenda is essential to assist institutions and teaching and learning centers in the development of training programs that cater to the need of the different types of populations of instructors.

Limitations

The exploratory nature of this study limits the generalizability of the results. Indeed, the sample size is small and only represents a particular slice of the STEM teaching professorate (i.e., physics and astronomy assistant professors). Thus, extrapolation to other STEM and non-STEM disciplines is not supported. Moreover, the participants in this study voluntarily chose to attend this pedagogical-focused workshop. Consequently, they may not represent typical new instructors in physics and astronomy. Finally, as reflective practice is inherently personal, it is possible that participants were not inclined to write about critical scenarios or to include controversial topics despite the confidential nature of this study.

Conclusion

This study is one of the first to provide an insight into the nature of STEM instructors' reflections on their teaching. The results show that physics and astronomy instructors with limited teaching experience are mostly unable to write reflections at a level that would promote pedagogical growth. This study thus points to the need to support and train STEM instructors on their reflective practices, especially if the intent of the inclusion of reflections in teaching evaluation processes is to promote instructional transformation.

Abbreviations
FATE   Framework for Assessing Teaching Effectiveness
STEM   Science, Technology, Engineering, and Mathematics
SET   Student Evaluation of Teaching

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s43031-024-00105-9.

Supplementary Material 1

Data availability
Due to the sensitivity of the data and to protect participants, we are unable to share data in any format with the public.

Declarations

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

References
Accelerating Systemic Change Network (2023). *Curated teaching evaluation initiative repository*. Retrieved 01/03 from https://ascnhighered.org/ASCN/evaluation_initiatives/repository.html.

Andrews, T. C., & Lemons, P. P. (2015). It's personal: Biology instructors prioritize personal evidence over empirical evidence in teaching decisions. *CBE—Life Sciences Education*, *14*(1), ar7. https://doi.org/10.1187/cbe.14-05-0084

Ansarin, A. A., Farrokhi, F., & Rahmani, M. (2015). Iranian EFL teachers' reflection levels: The role of gender, experience, and qualifications. *The Asian Journal of Applied Linguistics*, *2*(2), 140–155. https://caes.hku.hk/ajal/index.php/ajal/article/view/159.

Bain, J., Ballantyne, R., Mills, C., & Lester, N. (2002). *Reflecting on practice: Student teachers' perspectives*. Post Pressed.

Beane, R., McNeal, K. S., & Macdonald, R. H. (2019). Probing the National Geoscience Faculty Survey for reported use of practices that support inclusive learning environments in undergraduate courses. *Journal of Geoscience Education*, *67*(4), 427–445. https://doi.org/10.1080/10899995.2019.1621714.

Belvis, E., Pineda-Herrero, P., Armengol, C., & Moreno, V. (2012). Evaluation of reflective practice in teacher education. *European Journal of Teacher Education*, *35*(3), 1–14. https://doi.org/10.1080/02619768.2012.718758.

Betrabet Gulwadi, G. (2009). Using reflective journals in a sustainable design studio. *International Journal of Sustainability in Higher Education*, *10*(1), 43–53.

Boyer Commission on Educating Undergraduates in the Research University. (1998). *Reinventing undergraduate education: A blueprint for America's research universities*. State University of New York at Stony Brook for the Carnegie Foundation for the Advancement of Teaching.

Bradforth, S. E., Miller, E. R., Dichtel, W. R., Leibovich, A. K., Feig, A. L., Martin, J. D., Bjorkman, K. S., Schultz, Z. D., & Smith, T. L. (2015). University learning: Improve undergraduate science education. *Nature*, *523*(7560), 282–284. https://doi.org/10.1038/523282a.

Brookfield, S. D. (2017). *Becoming a critically reflective teacher*. Wiley.

Campoy, R. (2010). Reflective thinking and educational solutions: Clarifying what teacher educators are attempting to accomplish. *Srate Journal*, *19*(2), 15–22.

Chasteen, S. V., Perkins, K. K., Code, W. J., & Wieman, C. E. (2016). The science education initiative: An experiment in scaling up educational improvements in a research university. *Transforming Institutions: Undergraduate STEM Education for the 21st Century*, 125–139.

Day, C. (1993). Reflection: A necessary but not sufficient condition for professional development. *British Educational Research Journal*, *19*(1), 83–93.

Dennin, M., Schultz, Z. D., Feig, A., Finkelstein, N., Greenhoot, A. F., Hildreth, M., Leibovich, A. K., Martin, J. D., Moldwin, M. B., & O'Dowd, D. K. (2017). Aligning practice to policies: Changing the culture to recognize and reward teaching at research universities. *CBE—Life Sciences Education*, *16*(4), es5. https://doi.org/10.1187/cbe.17-02-0032.

Dewey, J. (1933). *How we think*. Courier Corporation.

Dinham, J., Choy, S. C., Williams, P., & Yim, J. S. C. (2021). Effective teaching and the role of reflective practices in the Malaysian and Australian education systems: A scoping review. *Asia-Pacific Journal of Teacher Education*, *49*(4), 435–449.

Dyment, J. E., & O'Connell, T. S. (2010). The quality of reflection in student journals: A review of limiting and enabling factors. *Innovative Higher Education*, *35*, 233–244. https://doi.org/10.1007/s10755-010-9143-y.

Farrell, T. S. (2003). *Reflective practice in action: 80 reflection breaks for busy teachers*. Corwin.

Fox, K. R., Campbell, M., & Hargrove, T. (2011). Examining reflective practice: Insights from pre-service teachers, in-service teachers and faculty. *Journal of Research in Education*, *21*(2), 37–54.

Gibbs, G. (1988). Learning by doing: A guide to teaching and learning methods. *Further Education Unit*.

Handal, G., & Lauvas, P. (1987). *Promoting reflective teaching: Supervision in practice*. Society for Research into Higher Education.

Hatfield, N., Brown, N., & Topaz, C. M. (2022). Do introductory courses disproportionately drive minoritized students out of STEM pathways? *PNAS nexus*, *1*(4), pgac167. https://doi.org/10.1093/pnasnexus/pgac167.

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, *48*(8), 952–984. https://doi.org/10.1002/tea.20439.

Huda, M., & Teh, K. S. M. (2018). Empowering professional and ethical competence on reflective teaching practice in digital era. In K. Dikilitas, E. Mede, & D. Atay (Eds.), *Mentorship Strategies in Teacher Education* (pp. 136–152). IGI Global. https://doi.org/10.4018/978-1-5225-4050-2.ch007.

Jay, J. K., & Johnson, K. L. (2002). Capturing complexity: A typology of reflective practice for teacher education. *Teaching and Teacher Education*, *18*(1), 73–85. https://doi.org/10.1016/S0742-051X(01)00051-8.

Jensen, S. K., & Joy, C. (2005). Exploring a model to evaluate levels of reflection in baccalaureate nursing students' journals. *Journal of Nursing Education*, *44*(3), 139–142. https://doi.org/10.3928/01484834-20050301-08.

Koester, B. P., Grom, G., & McKay, T. A. (2016). Patterns of gendered performance difference in introductory STEM courses. *arXiv preprint arXiv:1608.07565*.

Lane, A. K., Earl, B., Feola, S., Lewis, J. E., McAlpin, J. D., Mertens, K., Shadle, S. E., Skvoretz, J., Ziker, J. P., & Stains, M. (2022). Context and content of teaching conversations: Exploring how to promote sharing of innovative teaching knowledge between science faculty. *International Journal of STEM Education*, *9*(1), 1–16. https://doi.org/10.1186/s40594-022-00369-5

Larrivee, B. (2000). Transforming teaching practice: Becoming the critically reflective teacher. *Reflective Practice*, *1*(3), 293–306. https://doi.org/10.1080/713693162.

Larrivee, B. (2005). *Authentic classroom management: Creating a learning community and building reflective practice*. Allyn & Bacon.

Larrivee, B. (2008a). Development of a tool to assess teachers' level of reflective practice. *Reflective Practice*, *9*(3), 341–360. https://doi.org/10.1080/14623940802207451.

Larrivee, B. (2008b). Meeting the challenge of preparing reflective practitioners. *The New Educator*, *4*(2), 87–106. https://doi.org/10.1080/15476880802014132.

Lee, H. J. (2005). Understanding and assessing preservice teachers' reflective thinking. *Teaching and Teacher Education*, *21*(6), 699–715. https://doi.org/10.1016/j.tate.2005.05.007.

Lee, S. J. C., & Abdul Rabu, S. N. (2022). Google Docs for higher education: Evaluating online interaction and reflective writing using content analysis approach. *Education and Information Technologies*, *27*(3), 3651–3681. https://doi.org/10.1007/s10639-021-10760-4.

Loughran, J. J. (2002). *Developing reflective practice: Learning about teaching and learning through modelling*. Routledge.

Lubbe, W., & Botha, C. S. (2020). The dimensions of reflective practice: A teacher educator's and nurse educator's perspective. *Reflective Practice*, *21*(3), 287–300. https://doi.org/10.1080/14623943.2020.1738369.

Machost, H., & Stains, M. (2023). Reflective practices in education: A primer for practitioners. *CBE—Life Sciences Education*, *22*(2), es2. https://doi.org/10.1187/cbe.22-07-0148

Madsen, A., McKagan, S. B., & Sayre, E. C. (2017). Resource letter Rbai-1: Research-based assessment instruments in physics and astronomy. *American Journal of Physics*, *85*(4), 245–264. https://doi.org/10.1119/1.4977416.

Markkanen, P., Välimäki, M., Anttila, M., & Kuuskorpi, M. (2020). A reflective cycle: Understanding challenging situations in a school setting. *Educational Research*, *62*(1), 46–62. https://doi.org/10.1080/00131881.2020.1711790.

Marshall, T. (2019). The concept of reflection: A systematic review and thematic synthesis across professional contexts. *Reflective Practice*, *20*(3), 396–415. https://doi.org/10.1080/14623943.2019.1622520.

Marshman, E. M., Kalender, Z. Y., Nokes-Malach, T., Schunn, C., & Singh, C. (2018). Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm? *Physical Review Physics Education Research*, *14*(2), 020123. https://doi.org/10.1103/PhysRevPhysEducRes.14.020123.

Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., Weiner, B., & McKay, T. A. (2017). Patterns of gendered performance differences in large introductory courses at five research universities. *Aera Open*, *3*(4). https://doi.org/10.1177/2332858417743754.

McAlpine, L., & Weston, C. (2002). Reflection: Issues related to improving professors' teaching and students' learning. *Teacher Thinking Beliefs and Knowledge in Higher Education*, *59-78*. https://doi.org/10.1007/978-94-010-0593-7_4.

McAlpine, L., Weston, C., Berthiaume, D., Fairbank-Roch, G., & Owen, M. (2004). Reflection on teaching: Types and goals of reflection. *Educational Research and Evaluation*, *10*(4–6), 337–363. https://doi.org/10.1080/13803610512331383489.

Miller, E. R., & Fairweather, J. S. (2015). The role of cultural change in large-scale STEM reform: The experience of the AAU Undergraduate STEM Education Initiative. *Transforming Institutions: Undergraduate STEM Education for the 21st Century, 48*.

Minott, M. A. (2008). Valli's typology of reflection and the analysis of pre-service teachers' reflective journals. *Australian Journal of Teacher Education*, *33*(5), 55–65. https://doi.org/10.14221/ajte.2008v33n5.4.

Mohamed, M., Rashid, R. A., & Alqaryouti, M. H. (2022). Conceptualizing the complexity of reflective practice in education. *Frontiers in Psychology*, *13*, 1008234. https://doi.org/10.3389/fpsyg.2022.1008234.

Mooring, S. R., Mitchell, C. E., & Burrows, N. L. (2016). Evaluation of a flipped, large-enrollment organic chemistry course on student attitude and achievement. *Journal of Chemical Education*, *93*(12), 1972–1983. https://doi.org/10.1021/acs.jchemed.6b00367.

Munby, H., & Russell, T. (1989). Educating the reflective teacher: An essay review of two books by Donald Schon. *Journal of Curriculum Studies*, *21*(1), 71–80. https://doi.org/10.1080/0022027890210106.

O'Connell, T. S., & Dyment, J. E. (2004). Journals of post secondary outdoor recreation students: The results of a content analysis. *Journal of Adventure Education & Outdoor Learning*, *4*(2), 159–171. https://doi.org/10.1080/14729670485200511.

O'Connell, T. S., & Dyment, J. E. (2011). The case of reflective journals: Is the jury still out? *Reflective Practice*, *12*(1), 47–59. https://doi.org/10.1080/14623943.2011.541093.

Olson, S., & Riordan, D. G. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Report to the president*. Executive Office of the President.

Osterman, K. F., & Kottkamp, R. B. (2004). *Reflective practice for educators: Professional development to improve student learning*. Corwin.

Physics and Astronomy Faculty Teaching Institute (2023). *Faculty Teaching Institute, June 2023*. Retrieved 01/04 from https://www.physport.org/fti/.

Plack, M. M., Driscoll, M., Blissett, S., McKenna, R., & Plack, T. P. (2005). A method for assessing reflective journal writing. *Journal of Allied Health*, *34*(4), 199–208.

Pond, J. W., & Chini, J. J. (2017). Exploring student learning profiles in algebra-based studio physics: A person-centered approach. *Physical Review Physics Education Research*, *13*(1), 010119. https://doi.org/10.1103/PhysRevPhysEducRes.13.010119.

Rice, K. G., Lopez, F. G., & Richardson, C. M. (2013). Perfectionism and performance among STEM students. *Journal of Vocational Behavior*, *82*(2), 124–134. https://doi.org/10.1016/j.jvb.2012.12.002.

Richardson, G., & Maltby, H. (1995). Reflection-on-practice: Enhancing student learning. *Journal of Advanced Nursing*, *22*(2), 235–242. https://doi.org/10.1046/j.1365-2648.1995.22020235.x.

Ryan, M. (2013). The pedagogical balancing act: Teaching reflection in higher education. *Teaching in Higher Education*, *18*(2), 144–155. https://doi.org/10.1080/13562517.2012.694104.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.

Schön, D. A. (1991). *The reflective turn: Case studies in and on educational practice* (Vol. 131). Teachers College.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, *22*(2), 63–75.

Simonson, S. R., Earl, B., & Frary, M. (2022). Establishing a framework for assessing teaching effectiveness. *College Teaching*, *70*(2), 164–180. https://doi.org/10.1080/87567555.2021.1909528.

Sorby, S., Veurink, N., & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is 'yes'. *Learning and Individual Differences*, *67*, 209–222. https://doi.org/10.1016/j.lindif.2018.09.001.

Spalding, E., & Wilson, A. (2002). Demystifying reflection: A study of pedagogical strategies that encourage reflective journal writing. *Teachers College Record*, *104*(7), 1393–1421. https://doi.org/10.1111/1467-9620.00208.

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

Sumsion, J., & Fleet, A. (1996). Reflection: Can we assess it? Should we assess it? *Assessment & Evaluation in Higher Education*, *21*(2), 121–130. https://doi.org/10.1080/0260293960210202.

Tajeddin, Z., & Aghababazadeh, Y. (2018). Blog-mediated reflection for professional development: Exploring themes and criticality of L2 teachers' reflective practice. *TESL Canada Journal*, *35*(2), 26–50. https://doi.org/10.18806/tesl.v35i2.1289.

The University of Edinburgh (2021). *Reflecting on experience*. Retrieved 01/04 from https://www.ed.ac.uk/reflection/reflectors-toolkit/reflecting-on-experience.

The University of Kansas Center for Teaching Excellence. (2024). *Benchmarks for teaching effectiveness*. Retrieved 01/03 from https://cte.ku.edu/sites/cte/files/documents/programs-initiatives/KU%20Benchmarks%20Framework%202020update.pdf.

Thorpe, K. (2004). Reflective learning journals: From concept to practice. *Reflective Practice*, *5*(3), 327–343. https://doi.org/10.1080/1462394042000270655.

Valli, L. (1997). Listening to other voices: A description of teacher reflection in the United States. *Peabody Journal of Education*, *72*(1), 67–88. https://doi.org/10.1207/s15327930pje7201_4.

van Manen, M. (1977). Linking ways of knowing with ways of being practical. *Curriculum Inquiry*, *6*(3), 205–228. https://doi.org/10.1080/03626784.1977.11075533.

Weaver, G. C., Austin, A. E., Greenhoot, A. F., & Finkelstein, N. D. (2020). Establishing a better approach for evaluating teaching: The TEval Project. *Change: The Magazine of Higher Learning*, *52*(3), 25–31. https://doi.org/10.1080/00091383.2020.1745575.

Winchester, T. M., & Winchester, M. (2011). Exploring the impact of faculty reflection on weekly student evaluations of teaching. *International Journal for Academic Development*, *16*(2), 119–131. https://doi.org/10.1080/1360144X.2011.568679.

Wong, F. K., Kember, D., Chung, L. Y., & CertEd, L. Y. (1995). Assessing the level of student reflection from reflective journals. *Journal of Advanced Nursing*, *22*(1), 48–57. https://doi.org/10.1046/j.1365-2648.1995.22010048.x.

Wu, S. P., & Rau, M. A. (2019). How students learn content in science, technology, engineering, and mathematics (STEM) through drawing activities. *Educational Psychology Review*, *31*, 87–120.

Yik, B. J., Raker, J. R., Apkarian, N., Stains, M., Henderson, C., Dancy, M. H., & Johnson, E. (2022). Association of malleable factors with adoption of research-based instructional strategies in introductory chemistry, mathematics, and physics. *Frontiers in Education, 7*, 1016415. https://doi.org/10.3389/feduc.2022.1016415

Zahid, M., & Khanam, A. (2019). Effect of reflective teaching practices on the performance of prospective teachers. *Turkish Online Journal of Educational Technology*, *18*(1), 32–43.

Zeichner, K., & Liston, D. (1987). Teaching student teachers to reflect. *Harvard Educational Review*, *57*(1), 23–49. https://doi.org/10.17763/haer.57.1.j18v7162275t1w3w.

Publisher's Note