


RESEARCH

Open Access



# Characterizing faculty motivation to implement three-dimensional learning

Paul C. Nelson<sup>1\*</sup> , Rebecca L. Matz<sup>2</sup>, Kinsey Bain<sup>3</sup>, Cori L. Fata-Hartley<sup>1</sup> and Melanie M. Cooper<sup>1</sup>

## Abstract

The National Research Council's Framework for K-12 Science Education and the subsequent Next Generation Science Standards have provided a widespread common language for science education reform over the last decade. These efforts have naturally been targeted at the K-12 levels, but we have argued that the three dimensions outlined in these documents—scientific practices, disciplinary core ideas, and crosscutting concepts (together termed three-dimensional learning)—are also a productive route for reform in college-level science courses. However, how and why college-level faculty might be motivated to incorporate three-dimensional learning into their courses is not well understood. Here, we report a mixed-methods study of participants in an interdisciplinary professional development program designed to support faculty in developing assessments and instruction aligned with three-dimensional learning. One cohort of faculty ( $N=8$ ) was interviewed, and four cohorts of faculty ( $N=33$ ) were surveyed. Using expectancy-value theory as an organizational framework, we identified themes of perceived values and costs that participants discussed in implementing three-dimensional learning. Based on a cluster analysis of all survey participants' motivational profiles, we propose that these themes apply to the broader population of participants in this program. We recommend specific interventions to improve faculty motivation for implementing three-dimensional learning: emphasizing the utility value of three-dimensional learning in effecting positive learning gains for students; drawing connections between the dimensions of three-dimensional learning and faculty's disciplinary identities; highlighting scientific practices as a key leverage point for faculty ability beliefs; minimizing cognitive dissonance for faculty in understanding the similarities and differences between the three dimensions; focusing on assessment writing as a keystone professional development activity; and aligning local evaluation practices and promotion policies with the 3DL framework.

**Keywords** Expectancy-value theory, Motivation, Professional development, Three-dimensional learning, Undergraduate education

## Introduction

The National Research Council's Framework for K-12 Science Education (National Research Council, 2012a) precipitated a sharp acceleration in the science education community's "turn to practice" over the past decade

(Berland et al., 2016). Educators and researchers alike have benefitted from the emergent affordances of sharing a common language for describing and designing effective student learning experiences (Nehm, 2019). With a renewed emphasis on sensemaking and knowledge-in-use, along with a well-defined set of explicit signposts, three-dimensional learning (3DL) emphasizes the fusion of discipline-specific core ideas with the more generalizable crosscutting concepts and scientific practices.

Primary and secondary classrooms throughout the United States have implemented re-designed instruction, assessments, and even whole curricula based on these

\*Correspondence:

Paul C. Nelson  
nelso830@msu.edu

<sup>1</sup> Michigan State University, East Lansing, MI, USA

<sup>2</sup> University of Michigan, Ann Arbor, MI, USA

<sup>3</sup> Pepperdine University, Malibu, CA, USA

constructs as defined by the Framework and the subsequent Next Generation Science Standards (Kaldaras et al., 2021; Miller & Kastens, 2018; NGSS Lead States, 2013; Paprzycki et al., 2017). These are clearly non-trivial shifts; supporting transitions from students learning siloed facts to integrating the tools of science to explore and explain phenomena certainly demands a great deal from classrooms, teachers, schools, and districts (Anderson et al., 2018; McNeill et al., 2022). Perhaps unsurprisingly then, teacher uptake and implementation can look quite different at the individual level (Yang et al., 2020), which in turn facilitates rich strands of research geared toward answering questions about how 3DL can be most effectively introduced, operationalized, and situated in what are sometimes vastly different contexts.

Despite the generally warm embrace in the K-12 landscape (Gunckel & Tolbert, 2018; Rodriguez, 2015) and the potential for a more fluent and connected student learning trajectory, the adoption and adaptation of the language of 3DL among college science instructors and higher education researchers has been less apparent. Instead, a sustained focus on defining and characterizing “active learning” and its associated instructional practices and student outcomes has dominated discussions of college-level teaching and learning (Lombardi & Shipley, 2021). From one vantage point, this is a logical direction to steer toward, given that the typical undergraduate science classroom, especially at large universities, is still generally a didactic, content-covering lecture with little student engagement (Stains et al., 2018). Since those environments tend to yield unfavorable student outcomes (Freeman et al., 2014), it is certainly reasonable to try to understand the dynamics at play and work to support instructional change.

On the other hand, more superficial “plug and play” interventions (such as incorporating clickers) that are typically labeled as characteristic of active learning place too much emphasis on how content is delivered as opposed to what students are positioned to learn, integrate, and subsequently use. In fact, available evidence suggests that undergraduate students’ ability to use causal mechanistic reasoning to explain phenomena, for example, depends more on whether the course provides exposure to and practice with 3DL questions (in class meetings and on assessments) than on the presence of reformed instructional practices or the degree of in-class active learning (Ralph et al., 2022a, b).

Another persistent finding in undergraduate science education is the observation that knowing “what works” does not directly align with what actually happens in classrooms, even after years of convergent, triangulated research (Boyer, 1990; Seymour & Hewitt, 1997; Seymour & Hunter, 2019). An array of potentially interacting and

context-dependent potential explanations likely contributes to this observation. Similar to the situation at the K-12 level, this research-practice gap varies immensely across institutions, departments, and classrooms (Dancy & Henderson, 2008; Fairweather, 2008; Walczyk et al., 2007), but the drivers of such disparities are likely to differ in higher education in important ways. Institutional commitment to instructor development (Kezar & Eckel, 2002), paired with the ecology of local departmental culture (Reinholz & Apkarian, 2018), underpins diverse sets of teaching values and outcomes especially in research-intensive universities where instructors’ identities as teachers make up only a part of their overall personas (Brownell & Tanner, 2012; Zagallo et al., 2019).

One additional potentially limiting factor to change and uptake in the specific context of the current study is that the Framework (and thus 3DL) might be viewed as inappropriate or irrelevant for undergraduate-level content and learning goals. We have argued that this is not the case in general (Cooper et al., 2015) and that part of the translation to the college level involves building consensus among developers and instructors to define their disciplines’ core ideas (Lavery et al., 2016a), providing rich opportunities for building buy-in. While the K-12 scientific practices and crosscutting concepts map almost directly on to our recently-developed protocols used to characterize 3DL in undergraduate-level assessments (Lavery et al., 2016a) and instruction (Bain et al., 2020a), we contend that negotiating core ideas among faculty is necessary given the more specific nature of science coursework at the postsecondary level in broad contrast to K-12 (that is, college students take chemistry courses rather than physical science courses).

The purpose of the current study is to detail instructors’ motivation for implementing 3DL in order to establish further evidence for the relevance, feasibility, and potential impact of the Framework in the teaching and learning of college science. The literature basis describing science faculty motivation for engaging in professional development experiences is relatively small (McCourt et al., 2017; Wilson-Kennedy et al., 2019), and we are aware of no existing studies that investigate specifically the motivation of post-secondary faculty to implement 3DL. At the foundation of this analysis is the recognition that instructor buy-in and willingness to engage in adopting 3DL, on an individual level, is a necessary precondition and a proximate cause for generating opportunities for students to engage in 3DL in their undergraduate science courses. It is the depth and complexity of this shift (much more than a tweak in instructional practices or direct implementation of a prescribed curriculum) that likely drives the rich variety of responses that follow when instructors are asked to reflect on their experience.

## Theoretical framework

We used expectancy-value theory (EVT) as the key theoretical framework for characterizing instructor motivation in this study. Originally applied to educational contexts by Eccles and colleagues (1983) and extensively reviewed elsewhere (Wigfield & Cambria, 2010; Wigfield & Eccles, 2000; Wigfield et al., 2006), EVT describes motivation using the general constructs of value, expectancy, and cost. Value reflects how important a task is for helping someone reach their future goals, the extent to which they experience inherent enjoyment or pleasure in the task, and the importance of doing well on the task itself. Expectancy reflects the likelihood with which someone believes they can complete a task currently or in the future, both in absolute terms and compared to others. Cost reflects the effort, time, and resources required to engage in the task.

Generally, the EVT perspective assumes that a person's rational and individualistic motivation to engage in some target behavior can be described with two overarching constitutive components: the perceived value of the behavior (tempered by its cost), and the associated expectancy for success. If costs outweigh potential benefits, or if the actor perceives a low likelihood of success given their local context, EVT-defined motivation will be low. That is, the spirit of EVT is often captured with a mathematical model that assumes motivation ( $M$ ) can be described as a multiplicative relationship between expectancy ( $E$ ) and a "valence" value term, with costs ( $C$ ) subtracted from the perceived value ( $V$ ) of a given task:  $M = E(V - C)$ .

Certainly, EVT has been useful for describing choices and performance related to motivation in a wide range of educational contexts and with subjects of all ages, from adolescent children to college students and beyond. The general consensus from these studies (e.g., Wigfield et al., 2009) is that expectancies for success are more tightly coupled to student performance in a given domain, while task values are better direct predictors of course-taking intentions and choices. Expectancy and value are often positively associated with one another, leading to indirect effects that add predictive ability for various outcomes. Also as predicted by EVT, including perceived costs in the model accounts for additional variance in the parallel measurement of a range of student outcomes (e.g., Jiang et al., 2018).

Specifically with respect to STEM faculty, researchers have used EVT as a lens for understanding faculty motivation to engage in research-practice cycles (Matusovich et al., 2014; McPartlan et al., 2022), bridging the research-to-practice gap in implementing student-centered instruction (Finelli et al., 2014), and persisting with a long-term professional development experience

(McCourt et al., 2017). In the current study, we sought to characterize instructors' motivational profiles for adoption and persistent use of 3DL in their courses as defined by a variety of sub-constructs (e.g., utility value) under the broader umbrellas of each EVT dimension.

## Context

This study stems from a project initiated at a large, public doctoral university with very high research activity (The Carnegie Classification of Institutions of Higher Education, n.d.), herein referred to as RIU (for research-intensive university). This project also involved researchers and participants from three collaborating universities: two similar public doctoral universities with very high research activity and one public doctoral/professional university. With the support of two external grants and internal funding, a team of interdisciplinary researchers and practitioners across the sciences developed the project with the goal of transforming gateway introductory courses at each institution to be more coherently based on the principles of three-dimensional learning (Cooper et al., 2015; Matz et al., 2018a; National Research Council, 2012a).

A keystone activity within this project was the development and implementation of a professional development program, herein referred to as the Fellowship. The Fellowship is an intensive and extended program for faculty who regularly teach in large-enrollment courses (Fata-Hartley et al., 2023), akin to a faculty learning community (Cox, 2004; McCourt et al., 2017). The overall objective of the Fellowship is to support faculty in developing assessments and instruction that align with three-dimensional learning. Specifically, faculty identify the most pertinent disciplinary core ideas, crosscutting concepts, and practices for their courses (Cooper et al., 2017; Lavery & Caballero, 2018b; Cooper, 2020b); design, test, and reflect on assessments and instruction that elicit students' knowledge and abilities (Lavery et al., 2016a; Matz & Jardeleza, 2016b; Bain et al., 2020a); and over time become part of an interdisciplinary community committed to improving undergraduate education.

Extensive details about the Fellowship program are available in Fata-Hartley et al. (2023). Here, we summarize the key activities and approaches used to help Fellows achieve the target objectives. The earliest Fellowship meetings are focused on developing a sense of community while introducing 3DL. Readings, presentations, and discussions provide context and frame the research base for 3DL while examples and exemplars are explored. Fellows are introduced to the three-dimensional learning assessment protocol (3D-LAP; Lavery et al., 2016a) and the three-dimensional learning observation protocol (3D-LOP; Bain et al., 2020a)

as tools to develop and evaluate assessments and instructional activities, respectively, that incorporate core ideas, science practices, and crosscutting concepts. With this foundation, Fellows are charged with developing and implementing a teachable unit (Handelsman et al., 2007) with input and feedback from the Fellowship leaders as well as their peers in the program. A typical teachable unit includes a 3D learning objective, formative and summative assessments, and instructional materials such as class meeting plans and homework assignments. Fellows implement the unit, collect and analyze the assessment data and make refinements for future iterations. They present their teachable units and analyses as part of the effort to make connections across disciplines and build community.

The Fellowship has primarily drawn faculty from biology, chemistry, physics, and mathematics and statistics disciplines. This interdisciplinary nature of the Fellowship was purposeful, as such programs have been shown to contribute to broader educational change (Bouwma-Gearhart et al., 2014). Although the Framework and three-dimensional learning were originally developed for science and engineering, it was beneficial to involve mathematics and statistics faculty in the Fellowship as well given the pivotal place that these courses have in most STEM degree programs (McCormick & Lucas, 2011). Engineering was not part of the original local development team’s expertise nor the scope of the original grant, thus the engineering practices were not included.

Fellows were selected based on their level of experience teaching undergraduate courses, evidence of interest in and commitment to improving STEM education, evidence of potential impact on STEM education based on courses taught or targeted for reform, support from their chair or director, and potential for impact on departmental culture. Fellows participated in monthly meetings over 2 years and received a modest professional development budget to facilitate implementing 3DL in their

courses. The first and second cohorts were composed exclusively of faculty from RIU who taught introductory-level courses, while the third and fourth cohorts were expanded to include faculty from all four institutions as well as upper-division courses.

**Research questions**

We investigated the experiences of Fellows in the Fellowship and their motivation for engaging with 3DL so that we might better support future faculty engagement with 3DL and understand factors that impact how faculty engage with new teaching paradigms in general. In particular, the guiding research questions were: RQ1) What do Fellows perceive as supports and barriers as interpreted through expectancy-value theory that impact their motivation to integrate 3DL in their courses? and RQ2) How are these costs, values, and expectancies represented across the overall population of Fellows?

**Methods**

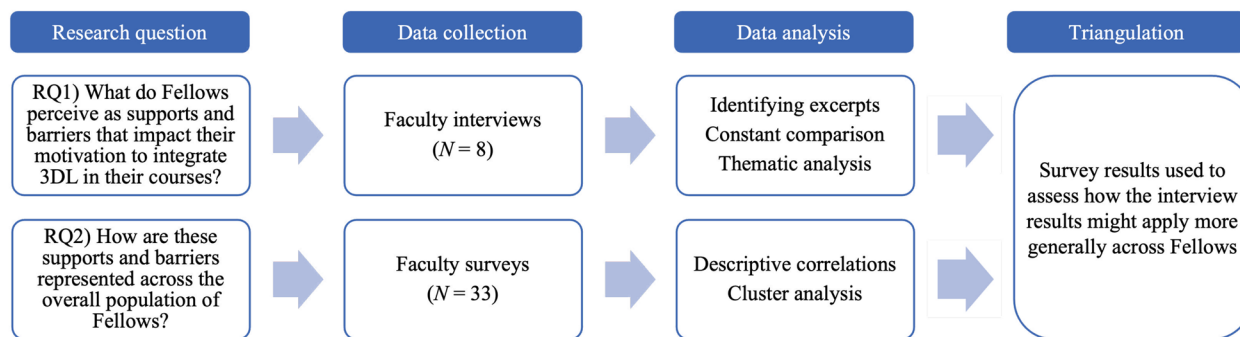
**Methodology**

In this convergent mixed-methods investigation (Creswell & Plano Clark, 2017), we conducted a qualitative thematic analysis (based on interview data) and a correlational quantitative study (based on survey data) for the purposes of comparing and contrasting results from each data set, leading to our overall interpretations (Fig. 1). In the analyses we focus on the qualitative data and use the survey results to provide insight into how the qualitative results might generalize to the broader population of Fellows. This study was determined by the Institutional Review Board at each university to be exempt from ongoing review.

**Interview study**

**Data collection**

Retrospective, semi-structured interviews were conducted with eight of nine possible Fellows from the first Fellowship cohort (Table 1), broadly aimed at exploring



**Fig. 1** Schematic showing the overall research design of the study

**Table 1** Characteristics of the interview participants

Pseudonym	Discipline	Employment track <sup>a</sup>	Years of teaching experience prior to Fellowship	DBER or SoTL focus prior to Fellowship? <sup>b</sup>	Part of the 3DL research team?
Abraham	Math & statistics	TT	13	Yes	No
Amanda	Chemistry	NTT	10	No	No
David	Physics	TT	1	Yes	Yes
Jim	Biology	NTT	11	Yes	Yes
Kari	Biology	TT	1	No	No
Kayla	Biology	TT	10	Yes	No
Lisa	Chemistry	TT	23	Yes	Yes
Scott	Physics	TT	16	No	Yes

<sup>a</sup> *TT* Tenure-track, *NTT* Non-tenure-track

<sup>b</sup> *DBER* Discipline-based education research, *SoTL* Scholarship of Teaching and Learning

instructors' conceptions of 3DL as well as supports and barriers that impacted if and how instructors implemented 3DL in their courses. We elected to interview a single cohort of Fellows completely local to RIU in an effort to limit the variation in contextual elements from one cohort to another. We attempted to interview the entire cohort because it was relatively small; only one Fellow did not respond to interview requests. Six participants were tenure-track faculty, and the two non-tenure-track participants were also long-term employees, having significant teaching experience and leadership roles in their respective departments.

Fellows' number of years as faculty prior to the Fellowship ranged from 1 to 23 years ( $M=11$  years). David and Kari were the two faculty newest to the RIU but had teaching experience as postdoctoral research associates and graduate student teaching assistants; all these Fellows thus had at least a few years of teaching experience when they began the Fellowship. Additionally, several Fellows had conducted research on teaching and learning and published in the realms of the scholarship of teaching and learning (Boyer, 1990) or discipline-based education research (National Research Council, 2012b) prior to their Fellowship term. We note this characteristic because we expected these Fellows might express different motivations for engaging with 3DL, given their greater familiarity with education research, than Fellows without. Finally, David and Lisa were involved as part of the 3DL research team even while they were Fellows. Jim and Scott joined the research team approximately 2 years after their Fellowship; thus, at the time that David, Jim, Lisa, and Scott were interviewed, all had had experience with 3DL beyond the Fellowship itself. Beyond being interviewed, none of these Fellows were privy to the theoretical framework or research project described in this paper at the time of the interview.

The semi-structured interview protocol (see Supporting Information 1) was developed to target supports and barriers for faculty change guided by the four frames model, an established framework for organizational culture and analysis that focuses on structures, symbols, people, and power (Bolman & Deal, 2017); we followed Reinholz and Apkarian (2018) who adapted this model in order to study change in STEM departments. The protocol was developed beginning from RQ1 and by writing complementary sub questions along with typical warm-up prompts (e.g., "What courses do you typically teach?") and comments to set the tone of the interview. The protocol was discussed amongst the research team, iteratively expanded and edited over several weeks, and piloted with David and Lisa.

Framing the interview as a general follow-up on their experience in the Fellowship, the remaining Fellows were recruited by email for interviews approximately 2 years after the end of their 2-year program. This timing was appropriate because we were interested in the Fellows' assessments of the extent to which they found 3DL attractive as an instructor having had several semesters since the Fellowship to modify their instruction and assessments (or revert to more traditional methods). Fellows were provided with a detailed month-by-month outline of the Fellowship activities to serve as a memory aide and were interviewed face-to-face in their offices, all by the same researcher (Author #3) who was then a postdoctoral research associate with the research team. Fellows were asked to review and sign an informed consent form and they were not compensated. The interviews were audio recorded and had durations from 39 to 91 min ( $M=64$  min). To protect participants as much as possible, we use pseudonyms and refrain from sharing any particularly sensitive or directly identifying data.

Following the warm-up questions about the instructor's teaching background and history at RIU, the

interview was organized into four sections, each with several open-ended questions: 1) motivation to join the Fellowship program and its perceived impact on the Fellow's personal teaching goals (e.g., "What motivated you to become a Fellow?"), 2) questions to elicit instructors' conceptual understanding of 3DL and its role in their courses (e.g., "What core ideas are most important to your courses and why?"), 3) questions to generate reflection on the factors supporting and preventing the implementation of 3DL (e.g., "What supports have you had in enacting 3D teaching practices?"), and 4) guided analysis of a set of discipline-specific assessment questions using the 3D-LAP (Laverty et al., 2016a) as a tool to direct the discussion (e.g., "How would you characterize these questions using the 3D-LAP?"). Fellows most often spoke about their motivations for implementing 3DL in the first three sections of the interview, and thus they provided the primary material for this analysis.

#### **Data analysis**

Two researchers (Author #1 and Author #2) conducted the main qualitative coding process. Author #1 was a postdoctoral research associate with the 3DL research team and a supporting facilitator for the then-current cohort of Fellows, designing 3DL-based activities, presentations, and meetings particularly with instructors of biology and mathematics courses. In both facilitation and coding, Author #1 drew on roughly a decade of math teaching experience and teacher professional development. Author #2 was a research specialist who had been involved with the 3DL research team and other STEM education projects across RIU for several years by that time. She was a chemistry instructor and brought some prior knowledge of Fellows' contexts, as well as university initiatives and projects related to 3DL (e.g., Matz & Jardeleza, 2016b).

Shortly after the interviews were conducted, the interviewer (Author #3) wrote a reflection on each interview in the form of an analytic memo (Saldana, 2021), representing the first dividing step between data collection and interpretation of the data. The audio recordings were transcribed by a third-party service and cleaned to correct speaker names and words, phrases, and acronyms specific to 3DL, the Fellowship program, and RIU. We used Dedoose to organize the transcripts, codes, and excerpts, ensuring that all excerpts retained unique information that could be used to trace back to the source transcript and audio files when broader context was needed for interpretation. Initially, we considered both the four frames model (Reinholz & Apkarian, 2018) and EVT frameworks as lenses to analyze the data. After an initial pass through the transcripts, we perceived that EVT would better support investigation of the research

question because the emerging ideas most clearly aligned with the goals, beliefs, perceptions, and identities of individuals rather than a larger unit like departments.

We developed an initial set of guidelines for coding the transcripts with EVT as the theoretical lens. With this structured approach (Corbin & Strauss, 2015), we drew categories and codes about expectancy and value from Eccles and Wigfield (1995) and Wigfield and Cambria (2010). Guided by Barron and Hulleman (2015) who formally introduced cost as a separate construct based on several studies showing that it loads as a separate motivational factor from expectancy and value (e.g., Kosovich et al., 2015), cost was considered its own distinct category. Specific codes for cost were drawn from Part et al. (2020), though we recognize that Eccles and colleagues described these dimensions of cost even in their original study where cost was a subcomponent of value (Eccles et al., 1983). Each code was described with a general definition, potential anchor examples generated from our initial analytic read, and proposed coding rules that applied the definition of the code to this study (see these coding guidelines and additional details in the Supporting Information 2).

Author #1 and Author #2 simultaneously but independently coded the same transcript according to the existing guidelines and then met to reconcile the code applications and note overall rules for the process (e.g., we allowed excerpts to be double-coded). With each transcript we iterated on and improved our shared understanding of the guidelines, using constant comparison to ensure that new excerpts were similar to those already coded (Lincoln & Guba, 1985). Both these authors continued to code and reconcile each transcript, thus all codes and excerpts reflect the consensus of these researchers.

This approach of "negotiated agreement" is one means of supporting the reliability and validity of results drawn from qualitative data; it is particularly appropriate when the goal of the study is to explore and generate new insights, or when greater sensitivity to subtle versus more overt meanings is desired (Campbell et al., 2013). Establishing reliability by negotiating agreement while coding qualitative data can be troublesome in the face of power differentials between the two coders (Campbell et al., 2013), such as between a research assistant and an advisor, but our two coders had no such differential; they were located in different units and had no reporting relationship during the time of the study. Both transitioned to new jobs during the time of data analysis and writing this paper with Author #2 moving to a separate university, providing some distance from the Fellowship environment and overall reducing their involvement in the activities of the broader research team.

Subsequently, themes that addressed the research questions were identified through finding repeated ideas as well as cutting and sorting (Ryan & Bernard, 2003). We note that the researchers who did the qualitative coding (Authors #1 and #2) collaborated with the researcher who conducted the interviews (Author #3) in periodic meetings and through early drafts of this manuscript to ensure that they had ample opportunity to provide feedback on the coding process and the interpretations of the data in light of their experience conducting the interviews. Excerpts have been lightly edited for grammar and vocal fillers to improve readability.

#### **Limitations of the interview data**

The interview study is limited in that we interviewed only one cohort of Fellows, all from RIU, meaning the perspectives of RIU Fellows are overrepresented relative to the overall Fellows population. This overrepresentation also colors our interpretations of the relationships between the interview and survey data and overall has led to a study that is essentially centered on experiences at RIU despite the Fellows program involving multiple institutions.

Additionally, the researchers who conducted the qualitative data analysis (Author #1 and Author #2) are different from the researcher who conducted the interviews (Author #3). While we all had reasonably good understanding of the Fellows program and the Fellows themselves, even the act of data collection is believed to begin the analytical process in qualitative studies (Saldana, 2021). Thus, our interpretations of the interview data may have missed subtle features of the interactions and aspects of the nonverbal communication that occurred during the interviews.

Further, the retrospective nature of the interviews was convenient and may have supported Fellows in contextualizing their dispositions with respect to 3DL and teaching with a long-term perspective (Sosniak, 2006). However, the time gap between the Fellowship and interviews also means that Fellows were interpreting their past selves, in particular their motivation for joining the Fellowship and implementing 3DL in their courses, through their then-present-day lens.

#### **Survey study**

##### **Data collection**

We also adapted a survey aligned with EVT to gauge Fellows' motivation for integrating 3DL in their courses (see Supporting Information 3). The survey items were modified from Eccles and Wigfield's (1995) large-scale study of perceived values and expectancies for success in secondary mathematics. We used this study as the

basis for our survey because their confirmatory factor analyses demonstrated that the EVT constructs were "clearly distinguishable" (p. 221) with both value and expectancy factors inversely related to cost factors. We added two items (Q21-22) to directly probe Fellows' motivation to continue using 3DL and to encourage others in their department to adopt 3DL. We asked Fellows about their perceived work balance between research and teaching as well as to describe teaching-related professional development programs in which they had recently participated. Finally, we asked them to select their discipline, cohort number, and whether they were in a tenure-track position.

We identified the sub constructs based on a combination of the EVT scales as presented by Eccles and Wigfield (1995) and the knowledge across our broader research team (over a dozen education researchers representing multiple institutions) of the teaching and learning context for post-secondary instructors in STEM courses. For example, none of the cost-related items presented by Eccles and Wigfield specifically referenced time as a constraint on motivation, but there is ample evidence that faculty consider time heavily as a cost in implementing evidence-based teaching practices (e.g., Matusovich et al., 2014). Similarly to McPartlan et al. (2022), we added two items toward this end under the sub-construct of opportunity cost (which was not explicit in the original survey). As another example, Eccles and Wigfield described the utility value of mathematics for students with respect to what they wanted to do after graduation. Our survey included a corollary statement considering the utility of 3DL for faculties' long-term career goals, but because we anticipated that 3DL might have additional utility value for instructors in terms of designing assessment questions we wrote an additional utility value-oriented item toward this end. In short, we modified the survey in what we viewed as an appropriate way, attempting to maintain fidelity to the spirit of the constructs as presented in the original survey while also speaking authentically to the faculty experience. Such expert review provides content validity evidence for the survey as it was implemented here (AERA et al., 2014).

The survey was built in Qualtrics using an end-anchored, numberless visual analog Likert-type scale for each item with 101 possible slider positions, recorded as 0 to 100 for analysis. The slider was initially centered at 50, and the anchors were specific to each item. Anonymous responses to the survey were solicited from the 55 Fellowship participants across all four cohorts and four institutions; individual links were sent out three times in an effort to maximize the overall response rate. At the time of the survey, Fellows

in Cohorts 1 and 2 had completed the Fellowship and were responding retrospectively; Fellows in Cohorts 3 and 4 were still in the midst of their Fellowship term.

### Data analysis

In total, 35 responses to the survey were collected from the 55 Fellows (64% response rate); two responses were then removed because they were incomplete. Though listwise deletion is often undesirable for handling missing data, the incomplete surveys represented a small proportion (6%) of the total collected and thus minimal bias was likely introduced (Baraldi & Enders, 2010). The remaining 33 responses used for analysis were roughly evenly distributed by discipline (6 biology; 10 chemistry; 7 math, statistics, or computation; 7 physics; and 3 other), cohort (6 Cohort 1, 10 Cohort 2, 11 Cohort 3, and 6 Cohort 4), and employment status (20 non-tenure-track and 13 tenure-track) based on survey questions 25, 26, and 27 (see Supporting Information 3), respectively. The proportions of respondents in each of these categories was not different from those in the overall population of Fellows (discipline:  $\chi^2(4) = 0.46$ ,  $p = 0.93$ ); cohort:  $\chi^2(3) = 0.25$ ,  $p = 0.97$ ); employment status:  $\chi^2(1) = 0.52$ ,  $p = 0.47$ ). Overall, 15 of the 55 Fellows were faculty at the three (non-RIU) collaborating institutions; thus, the majority of the survey results represent the perspectives of Fellows from RIU. Reliability coefficients for the survey by sub-construct are reported in Supporting Information 3.

Median values were used to summarize central tendencies of the instructor-pooled survey item response distributions; unless otherwise noted, a Wilcoxon rank-sum test  $p$ -value of less than 0.001 was required to justify any claim based on average differences across survey question responses. The full matrix is shown in Supporting Information 4. We used Kendall's  $\tau$  coefficient as a measure of the first-order associations between pairs of responses to individual survey items. Kendall's underlying non-parametric hypothesis test was chosen over those used to generate Pearson's or Spearman's coefficients because of its more robust performance with smaller sample sizes and non-normal response distributions (Schaeffer & Levitt, 1956). The resulting correlation matrix (Supporting Information 4) provides preliminary evidence of internal structure validity given that factor analyses were not possible (Knekta et al., 2019; see [Limitations of the survey data](#) section), and a way to visualize the relationships between expectancy, value, and cost items across our specific population.

As an exploratory method for describing Fellows' motivational profiles and assessing similarity between cohorts, their responses to the value, expectancy, and cost survey items were used as the basis for an unsupervised  $k$ -means cluster analysis. This analysis also included

the two survey items (Q21-22) that directly asked Fellows about their motivation for engaging with 3DL. We used clustering for the purpose of creating useful groups for describing the Fellows (Kaufman & Rousseeuw, 2009) rather than purporting that the clusters exactly uncover the "true" nature of their motivation; prior studies show that clustering has been successfully used for this purpose even with small sample sizes (Lane et al., 2021; Tasci, 2015; Zagallo et al., 2019) so long as the number of variables is smaller than the number of participants (Aldenderfer & Blashfield, 1984). We selected the  $k$ -means method in particular because among the more common clustering methods, it is less prone to problems when data overlap (Qu et al., 2007).

Fellows were partitioned into clusters using MATLAB (version 2019b) and the "cluster" method for choosing initial centroid positions. We ran several iterations of the analysis with different numbers of clusters; ultimately, we chose three clusters as the appropriate number because it provided maximum differentiation while still accounting for the relatively small sample size. To help ensure the clusters were robust, the convergent solution was identified following 1000 replications with different initial centroid positions. The final cluster solution reported here has the lowest total sum of point-to-centroid distances within each cluster.

### Limitations of the survey data

With respect to the survey study, three methodological limitations are salient.

First, given the small sample size, it was not possible to confirm that these data support the model of motivation that we specified with factor analysis. Typical sample size estimates given our parameters require a few hundred responses, but we have no way to increase the pool of Fellows so drastically. Having reached the upper-limit of  $N$  for this population, we yet maintain that the survey data offer valuable information relying on the evidence for content validity and the correlation matrix introduced earlier. We also note that an EVT-focused survey with similar items administered to several hundred biology faculty for their motivation to implement an evidence-based intervention showed a pattern broadly similar to our findings: higher value and expectancy and lower perceived costs all positively affect motivation (McPartlan et al., 2022). Reporting on the internal structure of similar surveys in similar populations is useful when stable estimates cannot be obtained with the given sample (Knekta et al., 2019).

Second, given that the four cohorts were run during different academic years, the survey was effectively administered at a different time point relative to the Fellowship for each cohort; that is, Cohorts 1 and 2 completed the



survey retrospectively, while Cohorts 3 and 4 completed the survey while still in the midst of Fellowship activities. We maintain that our alternative options—limiting the survey to a single cohort or waiting several more years to match the Fellowship-survey interval across all participants—would have resulted in a missed opportunity to share insights from these available data. Our finding of a broad similarity in terms of the relative frequencies of Fellows' motivational profiles across cohorts is consistent with the idea that the results reflect fundamental facets of the participants' experiences independent of when they were asked about them.

Third, the anonymous nature of the survey was purposeful because of the small sample size, still, this choice meant we could not reliably limit the survey responses to only those from RIU in order to remove the institutional variable in comparing results to those from the interviews.

## Results

In the following sections, we characterize the five most prominent emergent themes from the interview data and then use the survey responses to identify and describe trends in the broader population related to participants' motivation to engage with 3DL in the Fellowship and integrate 3DL into their courses.

### Interview study

#### *Theme 1: Fellows described 3DL as useful for engaging students, organizing their teaching, and transforming courses (utility value)*

Across all the interview data, utility value was the most commonly encountered EVT construct, with approximately one-third of coded excerpts falling into this category; all Fellows were represented by several utility value excerpts. In the majority of excerpts, Fellows described the utility of 3DL with respect to one or more of the following broad goals: 3DL was useful for helping students engage meaningfully with course content, 3DL was useful for communicating or organizing ideas about teaching, and 3DL was useful for developing or transforming a course or program.

The most referenced goal related to the utility value of 3DL was helping students gain a deeper understanding of course content. Lisa, for example, provided her perspective on the power of focusing on core ideas for students as they develop more refined disciplinary content knowledge:

*“Because the same core ideas keep coming up over and over again, the students get more comfort-*

*able thinking about the ideas, but I think they also develop a deeper understanding. Then when an idea comes up that's related, they can start making connections.”*

Lisa also identified the scientific practices as important levers that help prepare her students to use their knowledge later, saying,

*“I think the key benefit of trying to create situations for students to use the scientific practices is that it's actually more authentic to how they might be asked to use their understanding of chemistry subsequently ... It's preparing them better to actually become scientists. That's really our goal.”*

Aligned with the idea that the scientific practices are inherently engaging, David implied a benefit for so-called active learning, pointing out that in his hands, 3DL naturally results in active engagement: “If we engage them with three-dimensional learning, we fundamentally engage them actively. They have to be active. It's something they have to do, and they're going to learn the content.” Lisa further stated that, “In some ways, cross-cutting concepts provide a way of thinking about science in general,” suggesting that her views of the crosscutting concepts are less concrete but nonetheless convey optimism about their potential and importance for engaging students.

Perhaps as an intermediate step on the path to helping students gain a deeper understanding of science content, five instructors identified 3DL as a useful framework for effectively organizing and communicating their ideas about teaching. Jim, for example, emphasized the utility of 3DL in providing a framework for big ideas in biology, saying, “The Fellowship gave me a scaffold connected to my discipline that let me think about and organize the ideas I was teaching and why I was teaching them in a different way.” Similarly, Kayla said, “[3DL] really helped provide the language that I was needing to communicate both with my students and with my peers about the things I was trying to do in my class.”

Several instructors were even more specific about the formal organizational utility value of 3DL as they worked to transform courses or programs. Lisa summarized her work on developing a preparatory chemistry course, in particular, saying,

*“[The course] was intentionally designed around core ideas and scientific practices. The core ideas that I decided to focus on were electrostatic and bonding interactions, a little bit of structure and properties, and a little bit of energy. The three scientific practices that we explicitly work on are explanations, using models, and mathematics and computational*

*thinking.”*

In this work, Lisa limited her focus to a subset of core ideas and scientific practices, aligning with a common theme we observed that the core ideas and scientific practices take precedence over the crosscutting concepts in most instructors' minds when using 3DL to design curriculum and courses. David explicitly expressed this idea regarding a two-semester introductory physics sequence that he co-designed with colleagues, saying, “We are quite a bit more focused on 2D learning in [these courses]. There's a heavy emphasis on using your physics knowledge to engage in science practice.”

***Theme 2: Fellows found 3DL to be important because it aligns with their identity as scientists and teachers, as well as their ultimate purpose of training scientists (attainment value)***

We coded excerpts as reflecting attainment value when instructors conveyed that 3DL or engagement in the Fellowship was fundamentally important to them in some way. Jim, for example, shared a sense that 3DL was important because it offered a meaningful framework for teaching and learning that aligned with his identity as a scientist. In discussing the utility of 3DL as a scaffold for learning, he rhetorically asked the interviewer, “You want [students] to think like a scientist, right?” In alignment with prior work showing that post-undergraduate experiences are formative for faculty (Yoho et al., 2019), Jim continued to emphasize his identity as a scientist, saying, “I mean, I went to graduate school. I did a postdoc. That meant something to me.” Similarly, Jim said that he was never “really satisfied” with the model where students were memorizing a lot of detailed biological mechanisms. Rather, he said that on behalf of his students, “I wanted patterns. I wanted big ideas. I wanted it to make sense.” It was clearly important to Jim that students have appropriate tools for learning about biology as he did in his own scientific training.

Lisa felt that developing assessments aligned with 3DL was more time-consuming than developing traditional assessments, and she rationalized this choice by saying, “I [developed 3DL assessments] because I was interested and I thought it was important, but that didn't necessarily translate into acknowledgment [from administrators] that it was a valuable use of my time.” She went on to describe a sense of mission or purpose with respect to helping prepare students to “use their understanding of chemistry subsequently” with the scientific practices. Lisa said, “That's really our goal. We're training chemists. We'd like them to have

the habits and to think in the ways that scientists do.” Here, she conveyed the idea that “training chemists” is a central and valued aspect of her teaching mission and that 3DL provides an avenue towards this end.

***Theme 3: of the three dimensions, Fellows felt most capable of integrating scientific practices into their assessments and instruction (ability beliefs)***

The capacity to integrate scientific practices was a commonly referenced ability belief. For example, after Jim detailed how he uses the modeling practice to design instructional activities, he singled out the scientific practices when discussing the affordances of 3DL more generally saying, “One of the benefits [of 3DL] is some aspects of it are very intuitive to science instructors. Some of the practices, at least, conceptually resonate.” Kayla similarly found the practices to be the easiest entrée to 3DL for her, situating her perceived ability to integrate practices in relation to the core ideas in saying that “For me, it's easier to think in terms of practices than core ideas in general.” Amanda highlighted her facility with using and identifying the practices specifically when writing assessment questions:

*“The science practices, I think of more often. You know, what practice is incorporated in this question? Why would I ask them to do this? Is [this question] just a ‘memorize it back’ thing, which we obviously don't want to do all the time, or can I incorporate one of the practices?”*

In a related thread, Fellows drilled down further into the practices and emphasized their improved ability to elicit and probe student reasoning, a criterion required to reflect most of the scientific practices in the 3D-LAP and 3D-LOP protocols (Lavery et al., 2016a; Bain et al., 2020a). Following her experience with the Fellowship and subsequent involvement with the 3DL research team, Lisa reflected on her development on this front, saying, “I'm much more conscious of trying to draw out [students'] reasoning. I'm more likely to write questions that are going to ask ‘Why.’” Abraham, who in general had difficulty relating the practices to his content areas of mathematics and statistics, was nonetheless convinced that his assessments went beyond calculation to probe reasoning. He said,

*“All my assessments, I mean, they have computational aspects, but more importantly, emphasize what those computational numbers or answers mean. And I always ask students, what does this mean? What would it change if this is not true? We always give them assessment questions where they*

*have to reflect on their calculation.”*

As shown in Lisa’s and Abraham’s comments, these expressions of ability belief were stated predominantly in the context of writing assessment questions. Amanda provided another example with a window into how the emphasis on 3DL affects a typical chemistry assessment item, saying,

*“Before, we would ask [students], ‘Going across the row in the periodic table, what’s the trend in atomic radius?’ Or ‘Of these four elements, which one would be the smallest?’ Now, we may still ask them that question, but we will add on a second part, [asking for] the explanation. Don’t just memorize a trend, but why is the trend that way?”*

Presumably to prepare students for her revised assessment style and to align her instruction to the assessments, Amanda also noted, “It’s hard for me to remember what I did before, honestly. It’s so ingrained in me now that this is how we do it. Now, I definitely say ‘Why’ more in class.” Amanda viewed these concrete adjustments as part of a larger mindset shift, stating, “I think we treat [students] more like scientists now because we ask them to do things that scientists do.”

***Theme 4: Fellows’ barriers to implementing 3DL included developing their own understanding of the three dimensions, content coverage demands, and aligning curriculum across sections and courses (effort cost, opportunity cost)***

Fellows expressed several cost-related ideas as barriers to implementing 3DL in their courses. The ideas aligned both with effort cost, referring to the amount of effort required for success on a task or the demands incurred by engaging with a task, and with opportunity cost, which refers to the perceived amount of time required to be successful that could be used to engage with some valued alternative. This theme encompasses Fellows’ challenges in developing their own understanding of the three dimensions, negotiating changes to content coverage, and aligning curriculum across sections and courses.

Several Fellows voiced difficulties in operationalizing crosscutting concepts, core ideas, and, despite the clear trend of positive ability beliefs (Theme 3), scientific practices as well as in delineating between specific categories within a single dimension. Jim succinctly summarized these difficulties in integrating the language and definitions into his pedagogical practice in saying, “Sometimes when you try to get down to the nitty gritty of defining [the three dimensions], it gets very difficult.”

The crosscutting concepts were often perceived as the most nebulous dimension. For example, Lisa said, “The crosscutting concepts are really the hardest to get

a handle on ... They’re not all the same grain size. I think that’s one of the things that makes them difficult to really have a solid understanding of.” Scott and Jim voiced similar sentiments of frustration or uncertainty with respect to using the crosscutting concepts, as reflected when Jim said, “I’m still not quite sure what to do with the crosscutting concepts. Some days, I think I actually have a handle on how they’re useful. And then the next day, I’m not sure I do at all.”

Core ideas, although they were previously defined through consensus by groups of RIU faculty as disciplinary experts (Laverty et al., 2016a; Matz et al., 2018a), were not always consistent with individual instructors’ perspectives. Here, Jim reflected on the set of biology core ideas, saying, “I’m not convinced it’s a really good set of core ideas. To me, they’re not all there. They’re different from each other. Some of them don’t necessarily make a lot of sense to me.” He identified perceived shortcomings with respect to instruction around the specific core idea of systems in biology, saying that it is “a really important core idea” but that he’s “not sure we do a good job of actually teaching things from a systems perspective.”

The scientific practices and their associated definitions resonated more readily with the Fellows, except for Abraham as the sole mathematics and statistics instructor in the cohort. In fact, Abraham found all three dimensions to be dissonant or incommensurate with his home discipline of statistics, saying,

*“I struggled a lot to come up [with] ... these basic ideas, like, what is a scientific practice in statistics? I was struggling to see the difference between a scientific practice versus a core idea versus a crosscutting [concept]. When I see [them] for chemistry or biology, okay, I understand it. But when it came to statistics and putting them in those categories, I struggled a lot. I couldn’t see the differences. I couldn’t make a clear distinction ... I struggled a lot to be honest.”*

Although the Fellows were generally open to changing their instruction, several were transparent about the fact that the 3DL vision put forth in the Fellowship required significant effort to implement because it is largely incompatible with the familiar stand-and-deliver mode of university teaching. Fellows perceived that engaging students with 3DL in class meetings occurred at the expense of broader content coverage; notably, these were the only expressions of opportunity cost that explicitly identified the valued alternative. Scott made several comments towards this end throughout his interview, saying, for example:

*“The actual biggest effect [implementing 3DL] had was the amount of material that was covered. The*

*last time I taught an introductory course, I actually threw out pretty significant chunks of material to make more time to spend essentially on more careful and more layered clicker questions, trying to get more three-dimensional learning into the lecture. It takes time ... I think [I] actually ended up covering a little bit less, maybe 10% or 15% less."*

Later in the interview, Scott relayed the same idea but with an interesting twist where he commented on implementing 3DL as the version of himself that previously used more traditional, lecture-based teaching. He said,

*"The model of a professor jamming a bunch of information into a lecture ... that works against three-dimensional learning. And you feel an obligation -- 'Well, I have to cover angular momentum. These are the things I have to tell the students. I'm not doing my job if I don't spend time in class talking about them.' It's very hard to let go of that."*

Here, Scott gives a sense of how a traditional instructor—indeed, his former self—would justify the pedagogical choice to cover more material which Amanda paralleled in reflecting on instructors who rotate through the introductory chemistry courses, saying, "Sometimes they feel like they're giving up their time to tell information."

Fellows additionally expressed cost-related ideas about the implementation of 3DL beyond their particular course sections, sharing challenges associated with building coherent 3DL experiences for students across different sections of the same course and across multiple courses within a coordinated sequence. These efforts were most salient for Amanda and Jim as directors in their particular disciplines, overseeing coordination across several introductory courses. Amanda highlighted her difficulty in conveying the rationale for 3DL to faculty who rotated through the introductory chemistry courses, saying,

*"Getting them up to speed and getting them on board with three-dimensional learning is a challenge, and it's something that I haven't really tackled. It's challenging for me to explain to them why we're doing things the way we're doing."*

Similarly, Jim provided a glimpse into the nuances involved in his efforts to align the focus of assessments across different sections of a single course:

*"One of the challenges ... in trying to develop common assessments, is unless everybody has been exposed to and at least, on some level, understands and accepts the 3D framework as what you're going*

*for, if you put forward a common question that is three-dimensional, oftentimes it gets voted off the island, because other people don't understand what you're trying to do with it. They think it's too complicated or too convoluted or not the right thing to be doing."*

Jim experienced struggles in these kinds of negotiations with other instructors in part because they all had different levels of exposure to and acceptance of 3DL.

***Theme 5: Fellows described writing 3DL assessments as a key but costly activity in terms of effort and time (utility value, effort cost, opportunity cost)***

Within the paradigm of backward design and with the idea that students come to treat what is assessed as what is important, the Fellowship placed significant emphasis on the development of assessments that reflect 3DL. Throughout the interviews, we noticed that Fellows identified writing assessments as a key but costly leverage point for incorporating 3DL in their courses, and Fellows discussed the utility of both 3DL and the Fellowship towards this end. Jim, for example, oversees multiple introductory biology courses and discussed having additional teaching assistant (TA) time for grading constructed-response assessment items. He said that the Fellowship had provided him "a framework for what to do with those questions," that is, for writing better assessments and making good use of these TA resources. Lisa specifically cited the 3D-LAP (Laverty et al., 2016a) as useful with respect to writing questions involving calculations, saying:

*"If [students] did a numeric calculation, I'd always follow up with a question that made them use the numeric calculation. That was influenced by the development of the 3D-LAP and the idea that calculations in and of themselves are not as valuable to student learning. If you want to have students do a calculation, you should ask them to do something with the result."*

Lisa more broadly reflected on a specific article about scaffolding that the Fellows read, saying that she became "much more conscious" about how she writes assessment prompts and that she has carried this idea with her since the Fellowship, saying, "The whole idea of scaffolding has probably informed every exam question I've written since." Amanda agreed that "adding three-dimensional questions has helped" because, in contrast to traditional questions more focused on memorization, three-dimensional questions seem to more clearly reveal what students know. As a complement to the utility of 3DL, Kari

described the usefulness of Fellowship activities with respect to improving her assessments, saying, “Certainly, we spent a lot of time making better assessment items ... having group edits was very helpful.” Indeed, writing assessments as part of a team was a recurring theme among those who felt successful in adopting 3DL. Altogether, both 3DL itself and the Fellowship experience supported faculty in writing effective assessments for student learning, a key aspect of faculty work.

Still, several Fellows reflected on the difficulty and effort related to writing such assessments. Instructors stressed their struggles specifically in developing selected-response (in contrast to open-ended) questions, like Amanda who said, “I think it’s especially difficult to write multiple-choice questions that are three-dimensional. It’s much easier to write short answer questions.” Jim concurred that writing multiple-choice assessment items is challenging and admitted that sometimes he’d attempt to make an item 3D by “just throwing in a lot more words,” feeling unsatisfied with the results. Kari agreed that 3DL assessment items in general are “very hard to make” and in particular noted difficulties with linking different sub-parts of a question “without making it snowball or without giving away the answer.” Some instructors further found the convenience of using pre-existing assessments or question banks to be too tempting to override the difficulty of writing new 3DL questions. In alignment with our previous findings, particularly in physics (Matz et al., 2018a), Scott was open about his disappointment along these lines saying, “I took the easy way of [using pre-written] exams, where we have a big data bank of problems. So [the Fellowship] hardly changed my exams at all, sadly.”

Most prominent in opportunity cost, the time required to develop effective 3DL course materials, and particularly assessments, was expressed as a perceived limiting factor by four Fellows. Lisa explicitly identified this as one of the most significant barriers to high-fidelity implementation, saying:

*“Probably the biggest challenge when trying to do three-dimensional learning is developing materials for students to interact with ... developing assessments. I think all that takes more time than preparing worksheets for a traditional course. I think that’s one of the biggest barriers.”*

Amanda and Kari both echoed this sentiment in reflecting on challenges to implementation; Amanda said, “It takes more time to write the questions” and Kari said, “I allow more time now, when I write my exams, because I know how writing these in-depth questions take a lot of time.”

### Survey study

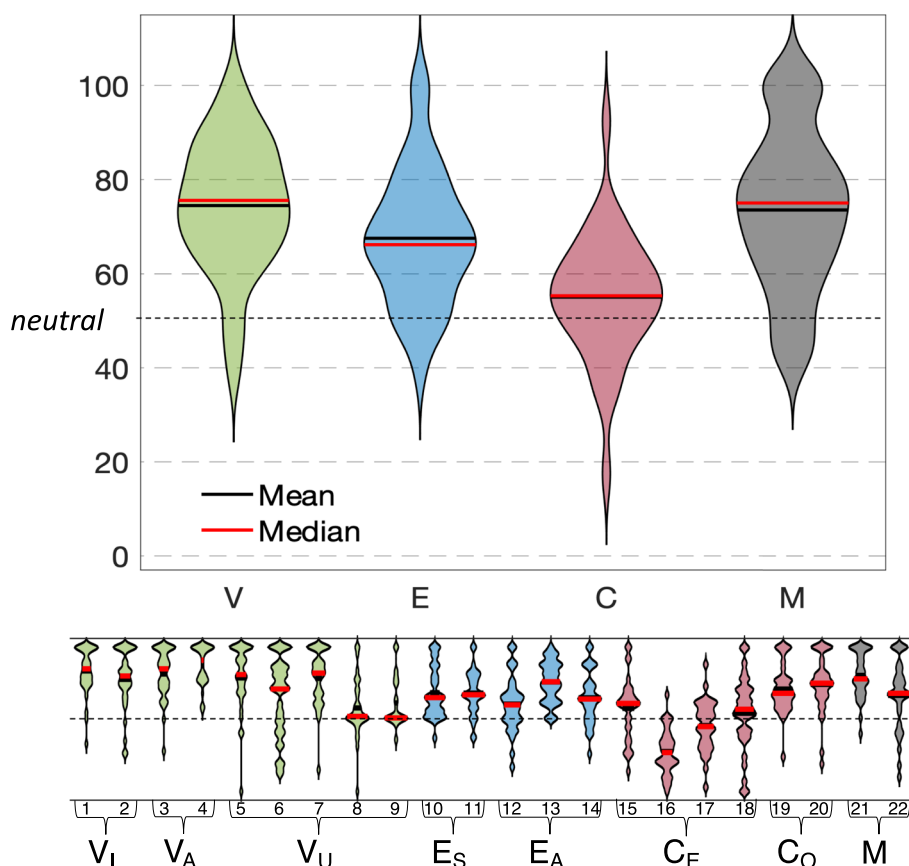
While the interview data provided a detailed look at the experiences of the first cohort of Fellows with respect to 3DL, the survey of faculty motivation for engaging with 3DL allowed for an examination of broader patterns across instructors and EVT constructs in ways that may be useful for refining future faculty development efforts.

### Overall patterns

If the EVT model  $M = E(V - C)$  holds, instructors with high motivation ( $M$ ) to implement 3DL in their courses will tend to have high expectancy ( $E$ ) and high perceived task value ( $V$ ) along with low associated costs ( $C$ ). Indeed, this general pattern emerged across the whole population of Fellows, assuming equal weighting for each within-construct item and averaging across instructors (Fig. 2); instructors assigned high value to 3DL relative to its perceived costs, with moderately positive expectancies for success. In the following subsections, we separately examine each of the main constructs and their associated survey item responses and then consider patterns across the individual instructors (interactions across constructs are described in the Supporting Information 4).

**Value** On average, Fellows reported intrinsic and attainment values associated with 3DL planning and implementation higher than any other EVT subconstruct with median response values of 80 or greater. Utility value items prompted more heterogeneous responses across participants and across specific survey questions; instructors tended to see 3DL as much more useful ( $p < 0.001$ ) in designing assessments ( $Mdn = 82$ ) and reaching long-term career goals ( $Mdn = 81$ ) than in supporting the more immediate review and promotion process ( $Mdn = 51$ ) or influencing student course evaluations ( $Mdn = 52$ ). The largest across-instructor spread of utility value prompt responses was found for the item assessing the usefulness of 3DL as a common language for discussing teaching with colleagues ( $Mdn = 71$ ); this is reflected in the bottom panel of Fig. 2 as a relatively thick lower tail of the violin plot for Q6.

Alongside high measures of central tendency, pairwise combinations of intrinsic and attainment value responses were strongly associated across instructors (see Supporting Information 4). The highest Kendall’s  $\tau$  magnitude among all items in the survey (0.73) was found between the first two items that focused on intrinsic value, quantifying levels of interest and enjoyment around planning 3DL instruction. The remaining combinations of the first four items were also all positively correlated ( $p < 0.005$ ).



**Fig. 2** Distributions of Fellows’ reported value (V), expectancy (E), cost (C), and motivation (M) on the survey. Points above 100 are an artifact of the violin plot smoothing function. Top: All questions aligned with each construct are averaged in the conglomerate response distributions. Bottom: Profiles for each individual question, grouped by intrinsic value (V), attainment value (V<sub>A</sub>), utility value (V<sub>U</sub>), expectancy for success (E<sub>S</sub>), ability beliefs (E<sub>A</sub>), effort cost (C<sub>E</sub>), opportunity cost (C<sub>O</sub>), and motivation (M)

In terms of utility value (Q5-Q9), the association patterns also showed positive correlations, both within the utility value category and extending to the intrinsic and attainment value subconstructs. Utility value items with higher median responses (Q5 and Q7) were most strongly associated with the responses to the intrinsic and attainment value components. Utility of 3DL as a common language and as an influencer of promotion and tenure processes was less predictive of attainment value and long-term career goal utility items.

*Expectancy* According to EVT, value alone is not enough; an actor must also believe that they can and will be successful. One way to classify the five items designed to probe instructor expectancy is to separate those that relate to the ultimate outcome of student success (Q10 and Q11) from those that focus on an instructor’s beliefs about their own 3DL teaching self-efficacy (Q12-Q14). In this data set, there was not a strong delineation in the responses to these two categories; the items elicited similar responses both in terms of measures of central

tendency (medians ranged from 60 to 76) and in their positive cross-instructor associations with one another ( $0.30 \leq \tau \leq 0.61$ ;  $p < 0.02$ ; see Supporting Information 4). The consistent and moderately positive magnitudes and associative relationships suggest that any one indicator could provide a first-order description of expectancy, and that factors presumably outside an instructor’s locus of control (implicitly included in Q10 and Q11) do not seem to have a strong impact on their perceived expectancy for success.

*Cost* Based on the interview data, it seems clear that sustained transformation of a course to support 3DL carries costs, both in terms of required effort and in the time required to develop materials and provide substantive feedback to students. Of course, instructors would be performing these tasks with or without 3DL; thus, most of the survey items pertaining to this construct framed effort or opportunity costs with respect to some reference (other instructors, other responsibilities, and traditional instruction or assessment). This framing clearly

had an impact on responses: median scores varied from 61 for Q15 (“In general, how hard is three-dimensional teaching for you?”) to 27 for Q16 (“Compared to most other instructors in your department, how hard is three-dimensional teaching for you?”), and there was no significant across-instructor association in the responses to this pair of items (see Supporting Information 4). When averaged across items (Q15-Q18), an overall difficulty or effort score of 48 suggests that Fellows do not see the work associated with 3DL as much harder than other teaching-related responsibilities. In contrast, the mean of the two time-focused items (Q19 and Q20) was higher (72;  $p < 0.001$ ) suggesting that the Fellows did perceive that preparing 3DL-aligned assessments and instruction takes longer than developing traditional curriculum materials and exams.

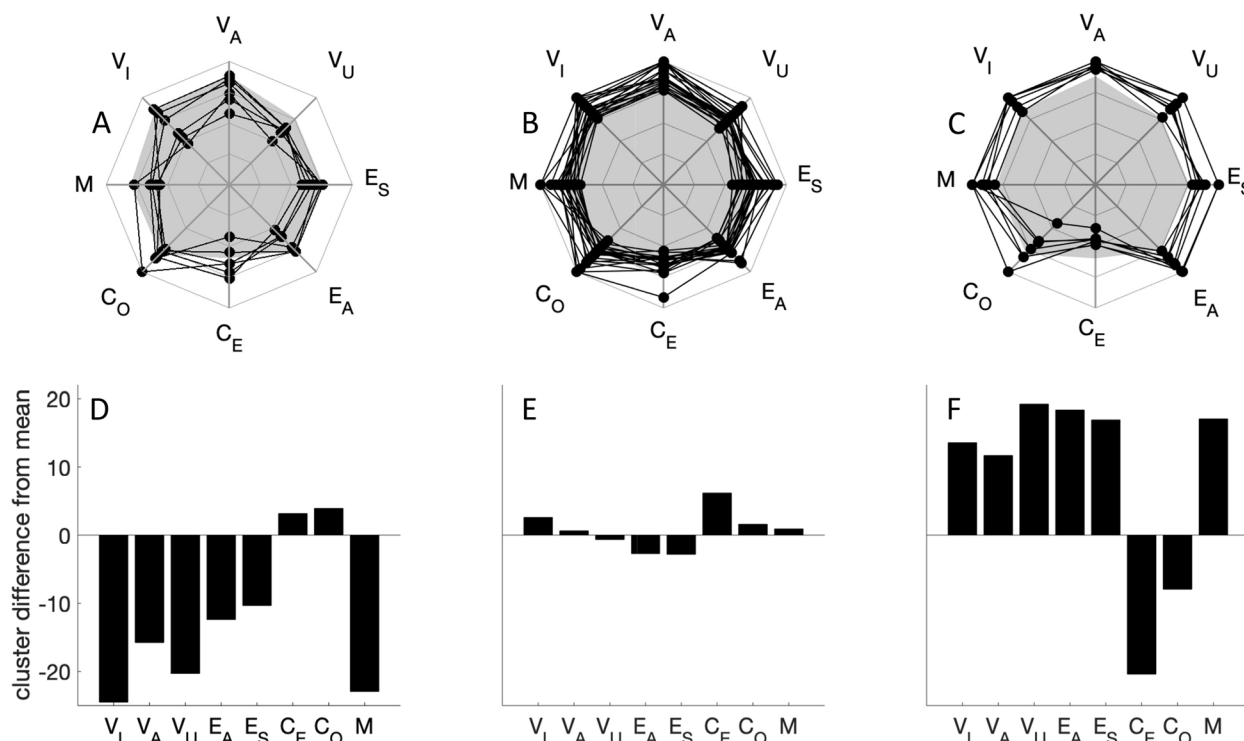
Relative to the value and expectancy patterns, Fellows’ reported perceptions of cost items were not as tightly coupled to one another: only 40% of the cost-cost survey item pairs were significantly associated at  $p < 0.05$  compared to 81% of value-value pairs and 100% of expectancy-expectancy pairs (see Supporting Information 4). In particular, responses to the time-related opportunity

cost items (Q19 and Q20) were largely independent of the effort- or difficulty-based item responses (Q15-Q18).

**Motivation** As an addendum to the modified Eccles and Wigfield (1995) survey items, we asked participants directly to gauge their motivation to use 3DL moving forward. Overall, Fellows tended to self-report somewhat higher levels of motivation ( $p = 0.016$ ) to continue using 3DL as a framework for their own instruction (Q21,  $Mdn = 78$ ) than they did to encourage others in their department to adopt 3DL (Q22,  $Mdn = 68$ ); these two metrics were significantly correlated ( $\tau = 0.40, p = 0.002$ ).

**Instructor profiles**

The overall survey response patterns described above provide one perspective of the Fellows’ EVT-based perceptions of 3DL based on measures of central tendency and across-instructor associations between pairs of survey item responses. Another approach is to investigate the potential existence and relative prevalence of common instructor profiles within this population; towards this end we applied an exploratory, unsupervised *k*-means clustering algorithm to the 33 sets of raw survey responses. To collapse and help visualize the



**Fig. 3** A-C Individual instructor profiles, grouped according to the three-cluster solution to a *k*-means procedure. D-F Corresponding deviations from grand-mean responses for each instructor cluster. Abbreviations: intrinsic value ( $V_I$ ), attainment value ( $V_A$ ), utility value ( $V_U$ ), ability beliefs ( $E_A$ ), expectancy for success ( $E_S$ ), effort cost ( $C_E$ ), opportunity cost ( $C_O$ ), and motivation ( $M$ )

resulting centroid-based classification, we used the eight variables defined for each individual instructor as the mean response of each EVT subconstruct (Q1-2: intrinsic value; Q3-4: attainment value; Q5-9: utility value; Q10-11: expectancy for success; Q12-14: ability beliefs; Q15-18: effort cost; Q19-20: opportunity cost; Q21-22: self-reported motivation). The across-instructor averages of these metrics are shown as the filled gray area of the radar charts in Fig. 3A-C; the 33 individual profiles are drawn on top of the average profile as solid black lines and profiles belonging to a given cluster are plotted together.

Some within-group similarities based on the eight-variable profiles emerged. The six instructors represented in Fig. 3A consistently had profiles with below-average value, expectancy, and self-reported motivation measures (falling inside the average response silhouette). This pattern is also shown in Fig. 3D where the cluster mean is compared to the full-population mean for the eight subconstructs: negative differences reflect below average responses. The group was much less distinguishable from the mean based on its reported perceptions of the costs connected to 3DL, which is at least intuitively consistent with the relatively low associations between cost-based items and others described in the “Overall patterns” section.

More than half (20 of 33) of the participants generated survey responses that fell into a large cluster that can be described as typical and perhaps unremarkable. These individual profiles are shown in Fig. 3B along with the collapsed difference-from-mean summaries in Fig. 3E. Given the relatively high median reports of value across the whole population (an average of 75 across Q1-9), expectancy (66), and self-reported motivation (73), compared to that of costs (56), one could reasonably predict that most of the Fellows from this group would intend to continue adopting and adapting the 3DL framework in their courses moving forward.

The seven individuals sorted into the third group of Fellows (Fig. 3C and F) were striking because of their notable deviations from average responses for all eight subconstructs. Except for opportunity cost (which is a poor predictor of value, expectancy, and motivation; see Supporting Information 4), this cluster’s collapsed profile is more than 10 units away from the average for every EVT category, and always in the direction that would suggest stronger predicted motivation based on the  $M = E(V - C)$  model. In other words, these seven Fellows had relatively high value, expectancy, and self-reported motivation, along with comparatively low perceived costs associated with 3DL. At a minimum, their responses suggest that their externally presented

values and attitudes toward 3DL would not be a barrier to adoption.

**Cluster distributions**

The key question we aimed to address with the survey data is whether the first cohort of Fellows (from which comprised the entire population of interview participants in this study was drawn) was representative of later groups or unique in its motivational profile composition. The distribution of individual response profiles based on cluster type and disaggregated by cohort number is shown in Table 2. A few points are worth noting regarding the way cluster types were associated with various descriptive features of the instructors.

Cohort 1 did not stand out in terms of its cluster composition: response types from each cluster were included and represented at a rate consistent with the overall population. Cohort 4 did not include any representatives from Cluster C (higher motivation). Cluster A responses (lower motivation) were spread across all four cohorts also in a way roughly proportional to the size of each cohort. Overall, based on cluster counts, it appears that Cohort 1 is a reasonable proxy for later cohorts and the distribution of the more extreme clusters (A and C) does not vary strongly across cohorts (Fisher-Freeman-Halton Exact Test  $p=0.90$ ).

Additionally, the three cluster profiles were observed as roughly proportional to the group size when disaggregated by discipline (biology, chemistry, mathematics/statistics, and physics), tenure status (tenure stream vs. non-tenure stream), and the instructors’ reported balance between research and teaching (Q23; research-focused vs. teaching-focused). None of the Fisher-Freeman-Halton Exact Tests designed to identify expectancy outliers resulted in  $p$ -values lower than 0.3, indicating that the prevalence of each cluster type does not depend on these more objective classifications (raw count tables not shown). It should be noted that our relatively small sample size precludes strong generalization of this essentially negative result.

**Table 2** Survey response cluster profile counts for four cohorts of Fellows

Cohort	Cluster			Total
	A	B	C	
1	1	4	1	6
2	2	5	3	10
3	2	6	3	11
4	1	5	0	6
<b>Total</b>	<b>6</b>	<b>20</b>	<b>7</b>	<b>33</b>



## Discussion

The principal findings from this study are the perceived values and costs we identified that impacted college instructors' motivation to integrate 3DL in the courses that they teach. Some value sub-constructs, like the intrinsic value associated with using 3DL to help students develop deeper disciplinary understanding, were reported to support ongoing motivation in both the interview and survey data. This finding is encouraging, as intrinsic and autonomously regulated sources of motivation are more likely to lead to sustained behavioral change than those based on controlled sets of rewards, punishment, and compliance or the approval of others (Ryan & Deci, 2000). In fact, incorporating 3DL had little effect on more externally regulated potential sources of utility value such as teaching evaluations and promotion processes. While this incongruence might be a source of frustration for faculty development facilitators and those interested in pedagogical change more broadly, it could also be seen as a clear opportunity for administrators to better align reward systems with the values of effective and dedicated instructors (Aster et al., 2021; Sansom et al., 2023). Tightening the alignment between institutional instructor evaluation and a framework like 3DL is unlikely without evidence of efficacy and feasibility in relevant contexts; this study is part of a larger narrative aimed at laying such a foundation of evidence and reasoning as rationale for such a shift.

As with any significant behavioral change, negative trade-offs are likely to surface (in the form of costs in the language of EVT), and such emergent costs are ignored at a change agent's peril. Our data suggest that the additional time requirements are more salient costs for potential adopters as compared to the effort required or difficulty of implementation. The perception of time scarcity for changing teaching practices is a common (if not particularly surprising) barrier that has also been noted elsewhere (e.g., Finelli et al., 2014; McCourt et al., 2017; Sansom et al., 2023). To minimize perceived opportunity costs, the Fellowship has evolved over time to emphasize adaptation, growth, and "trying something" over rigid blanket adoption or an implication that every class activity and assessment question should be three-dimensional. Anecdotally, but also in line with the interview data presented here, even Fellows with relatively traditional teaching philosophies found it useful to regularly refer to the scientific practices when designing assessments. Because the Fellowship also emphasizes backward design (Fata-Hartley et al., 2023), this small and low-cost behavior has the potential to act as a nudge toward significant and meaningful changes in instruction and student learning.

Even though the overall set of Fellows is relatively small, we found value in categorizing them into different motivational profiles based on their survey responses as an initial step in identifying categories of faculty that might benefit from different professional development experiences (Bae et al., 2020). Larger professional development programs with more teachers or faculty may find such characterizations even more useful and perhaps identify subcategories of the three profiles reported herein. Future research should also be designed to more directly determine whether any "pre-existing" instructor characteristics can reliably predict motivational profile types, and how the self-reported measures of the survey map on to actual implementation of 3DL assessment and instruction. It also remains an open question as to how these types of measures change over time, both within individuals and with respect to more precise timing of the data collection relative to Fellows' participation in the program.

An important second-order finding of the current study is that the distribution of motivational profiles among the Fellows that were interviewed appears to generalize across the broader population of Fellows, providing an encouraging indication that the interview data might also reasonably speak to the experiences of the other cohorts. This was somewhat surprising, as many of the "early adopters" of the first cohort have since become integral members of the local 3DL research team, and one might have reasonably expected starker differences in the relative proportions of the different profiles. With the small sample size tempering the exuberance of our interpretation, the results are at least consistent with the idea that the "ingredients" of the Fellowship appear to be more important than the "cooks" of the first cohort. At the same time, we acknowledge the self-selected nature of all the Fellowship participants. In describing faculty motivation for implementing 3DL, we have only included the perspectives of faculty who were already ostensibly interested in 3DL, or at minimum willing and able to commit to a significant professional development experience. The specific values, costs, and expectancies for a more widely representative group of faculty may well be different and thus indicate different potential approaches to professional development.

## Limitations

Additional limitations merit discussion. For example, the model used in calculating the overall patterns for motivation based on the survey data ascribes equal weight both for the items within a given construct (i.e., every value item has the same weight) and for the value, expectancy, and cost categories themselves. While this may be a typical approach for studies using EVT, whether Fellows

experience these constructs in equal proportion with respect to their motivation for implementing 3DL is an open question. Beyond this, Stolk et al. (2021) observed the motivation of college STEM students across almost a dozen institutions to change over the course of a term, and it is reasonable to assume that STEM faculty motivation might be similarly dynamic.

Additionally, throughout this article we have described “implementing 3DL” somewhat simplistically as though there is a binary characterization of courses as reflecting 3DL or not. Certainly, the continuum of uptake and adoption is more complicated, shown even in our own work where we observed ranges from section to section in the proportion of exam points allocated to questions that have the potential to elicit evidence of 3DL (Matz et al., 2018a). The participants in the current study were each at unique points in their ongoing process of conceptualizing, adapting, and integrating 3DL in their courses. This was undoubtedly associated with varied levels of competence in transforming and aligning assessments and class activities with the normative criteria that are laid out and operationalized in our protocols. A clear direction for future work would be to compare objectively measured 3DL uptake of individual instructors (via assessments and class observations) to their reported perceptions and intentions reflected in the interview and survey data.

### Implications and conclusions

With an awareness that any change effort needs to “know” the local context (Lund & Stains, 2015; Shadle et al., 2017), the key implications of this study are general in the sense that they are likely to be relevant for a diverse population of college-level science instructors, and specific in that they are framed around the foundations of building expectancies for success and task value while reducing perceived costs of 3DL. Based on the emergent EVT-filtered interview themes and survey data, we recommend the following to improve faculty motivation for implementing 3DL: emphasizing the utility value of 3DL in effecting positive learning gains for students (based on Theme 1); drawing connections between the dimensions of 3DL and faculty’s disciplinary identities (Theme 2 and the attainment and intrinsic value survey items); highlighting scientific practices as a key leverage point for faculty ability beliefs (Theme 3); minimizing cognitive dissonance for faculty in understanding the similarities and differences between the three dimensions (Theme 4); focusing on 3D assessment development as a keystone course transformation activity (Theme 5); and aligning local evaluation practices and promotion policies with the 3DL framework (utility value survey items).

Given that teacher motivation can positively influence student interest and achievement (Keller et al., 2017), supporting faculty motivation for engaging with 3DL in their courses may well lead to gains in cognitive and affective student outcomes in college-level science courses (e.g., Ralph et al., 2022a, b), and perhaps STEM courses more broadly. In parallel with attempts to understand individual instructors’ perceptions of 3DL, it is also critical to consider the situated context of their roles within departments (Reinholz & Apkarian, 2018) and institutions (Kezar & Eckel, 2002). It is only against the backdrop of this broader, layered landscape that students will ultimately have the potential opportunity to experience 3DL consistently across disciplines and over time. This study adds to the growing and convergent body of evidence pointing to 3DL as a powerful framework and language to support the development of meaningful knowledge-in-use, both in K-12 settings and now in higher education. What are institutions willing to commit to support and value the type of teaching and learning that we need to put this knowledge into use?

### Abbreviations

3DL	Three-dimensional learning
EVT	Expectancy-value theory
M	Motivation
E	Expectancy
V	Value
C	Cost
RIU	Research-intensive university
TA	Teaching assistant

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s43031-023-00079-0>.

**Additional file 1.** Interview protocol.

**Additional file 2.** Interview coding details and guidelines.

**Additional file 3.** Survey protocol.

**Additional file 4.** First-order survey response correlations and patterns across constructs.

### Acknowledgements

This work was made possible by the generous participation of the Fellows, and was further supported by Abigail I. Green, who helped clean the interview transcripts, and the 3DL research team.

### Authors’ contributions

PN and RM analyzed and interpreted the interview and survey data and contributed equally to writing the manuscript. PN developed the survey and collected the associated data. KB developed the interview protocol and collected the associated data. CFH and MC facilitated the fellowship program. All authors contributed to writing and editing and approved the final manuscript.

### Funding

Research funding was provided by the Association of American Universities’ Undergraduate STEM Education Initiative (funded by the Helmsley Charitable Trust), the Office of the Provost at Michigan State University, and the National

Science Foundation under #1725395, #1725520, #1725609, and #1726360. These funding bodies played no role in the design of the study; the collection, analysis, and interpretation of the data; or the writing of the manuscript.

#### Availability of data and materials

The data that support the findings of this study are not openly available due to reasons of sensitivity. Requests should be directed to the corresponding author.

#### Declarations

#### Competing interests

The authors declare that they have no competing interests.

Received: 12 May 2023 Accepted: 31 July 2023

Published online: 14 August 2023

#### References

- Aldenderfer, M., & Blashfield, R. (1984). *Cluster analysis*. Sage Publications.
- American Educational Research Association, American Psychological Association, and National Council for Measurement in Education (AERA, APA, and NCME). (2014). *Standards for educational and psychological testing*.
- Anderson, C. W., de los Santos, E. X., Bodbyl, S., Covitt, B. A., Edwards, K. D., Hancock, J. B., II, Lin, Q., Morrison, C., Penuel, W. R., & Welch, M. M. (2018). Designing educational systems to support enactment of the Next Generation Science Standards. *Journal of Research in Science Teaching*, 55(7), 1026–1052. <https://doi.org/10.1002/tea.21484>
- Aster, E. M., Bouwma-Gearhart, J., & Fisher, K. Q. (2021). Contextualizing communities in an instructional improvement initiative: Exploring STEM faculty engagement in teaching-related conversations. *Disciplinary and Interdisciplinary Science Education Research*, 3(1), 1–22. <https://doi.org/10.1186/s43031-021-00038-7>
- Bae, C. L., Hayes, K. N., & DeBusk-Lane, M. (2020). Profiles of middle school science teachers: Accounting for cognitive and motivational characteristics. *Journal of Research in Science Teaching*, 57(6), 911–942. <https://doi.org/10.1002/tea.21617>
- Bain, K., Bender, L., Bergeron, P., Caballero, M. D., Carmel, J. H., Duffy, E. M., Ebert-May, D., Fata-Hartley, C. L., Herrington, D. G., Laverty, J. T., Matz, R. L., Nelson, P. C., Posey, L. A., Stoltzfus, J. R., Stowe, R. L., Sweeder, R. D., Tessmer, S. H., Underwood, S. M., Urban-Lurain, M., & Cooper, M. M. (2020a). Characterizing college science instruction: The three-dimensional learning observation protocol. *PLOS ONE*, 15(6). <https://doi.org/10.1371/journal.pone.0234640>
- Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 48(1), 5–37. <https://doi.org/10.1016/j.jsp.2009.10.001>
- Barron, K. E., & Hulleman, C. S. (2015). Expectancy-value-cost model of motivation. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., Vol. 8, pp. 503–509). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.26099-6>
- Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*, 53(7), 1082–1112. <https://doi.org/10.1002/tea.21257>
- Bolman, L. G., & Deal, T. E. (2017). *Reframing organizations: Artistry, choice, and leadership*. Wiley.
- Bouwma-Gearhart, J., Perry, K. H., & Presley, J. B. (2014). Improving postsecondary STEM education: Strategies for successful interdisciplinary collaborations and brokering engagement with education research and theory. *Journal of College Science Teaching*, 44(1), 40–47.
- Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Princeton University Press. <http://eric.ed.gov/?id=ED326149>
- 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>
- Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3), 294–320. <https://doi.org/10.1177/0049124113500475>
- Cooper, M. M., Caballero, M. D., Ebert-May, D., Fata-Hartley, C. L., Jardeleza, S. E., Krajcik, J. S., Laverty, J. T., Matz, R. L., Posey, L. A., & Underwood, S. M. (2015). Challenge faculty to Transform Stem Learning. *Science*, 350(6258), 281–282. <https://doi.org/10.1126/science.aab0933>
- Cooper, M. M., Posey, L. A., & Underwood, S. M. (2017). Core ideas and topics: Building up or drilling down? *Journal of Chemical Education*, 94(5), 541–548. <https://doi.org/10.1021/acs.jchemed.6b00900>
- Cooper, M. M. (2020b). The crosscutting concepts: Critical component or “third wheel” of three-dimensional learning? *Journal of Chemical Education*, 97(4), 903–909. <https://doi.org/10.1021/acs.jchemed.9b01134>
- Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. SAGE Publications.
- Cox, M. D. (2004). Introduction to faculty learning communities. *New Directions for Teaching and Learning*, 2004(97), 5–23. <https://doi.org/10.1002/tl.129>
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Dancy, M., & Henderson, C. (2008). Barriers and promises in STEM reform. In *Commissioned paper for National Academies of Science Workshop on Linking Evidence and Promising Practices in STEM Undergraduate Education*.
- Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents’ achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, 21(3), 215–225. <https://doi.org/10.1177/0146167295213003>
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., & Meece, J. L. (1983). Expectations, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). W.H. Freeman.
- Fairweather, J. (2008). *Linking evidence and promising practices in Science, Technology, Engineering, and Mathematics (STEM) undergraduate education [A status report for the National Academies National Research Council Board of Science Education]*.
- Fata-Hartley, C. L., Nelson, P. C., Cooper, M. M., & Matz, R. L. (2023). Changing How STEM Faculty Teach by Changing What They Teach. In S. Linder, C. Lee, S. K. Stefl, & K. A. High (Eds.), *Handbook of STEM faculty development*. Information Age Publishing, Inc.
- Finelli, C. J., Daly, S. R., & Richardson, K. M. (2014). Bridging the research-to-practice gap: Designing an institutional change plan using local evidence. *Journal of Engineering Education*, 103(2), 331–361. <https://doi.org/10.1002/jee.20042>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafo, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111, 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Gunckel, K. L., & Tolbert, S. (2018). The imperative to move toward a dimension of care in engineering education. *Journal of Research in Science Teaching*, 55(7), 938–961. <https://doi.org/10.1002/tea.21458>
- Handelsman, J., Miller, S., & Pfund, C. (2007). *Scientific teaching*. Macmillan.
- Jiang, Y., Rosenzweig, E. Q., & Gaspard, H. (2018). An expectancy-value-cost approach in predicting adolescent students’ academic motivation and achievement. *Contemporary Educational Psychology*, 54, 139–152. <https://doi.org/10.1016/j.cedpsych.2018.06.005>
- Kaldaras, L., Akaeze, H., & Krajcik, J. (2021). Developing and validating Next Generation Science Standards-aligned learning progression to track three-dimensional learning of electrical interactions in high school physical science. *Journal of Research in Science Teaching*, 58(4), 589–618. <https://doi.org/10.1002/tea.21672>
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*. Wiley.
- Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers’ pedagogical content knowledge and motivation on students’ achievement and interest. *Journal of Research in Science Teaching*, 54(5), 586–614. <https://doi.org/10.1002/tea.21378>
- Kezar, A., & Eckel, P. D. (2002). The effect of institutional culture on change strategies in higher education. *The Journal of Higher Education*, 73(4), 435–460. <https://doi.org/10.1080/00221546.2002.11771159>

- Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE Life Sciences Education*, 18(1), rm1. <https://doi.org/10.1187/cbe.18-04-0064>
- Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in middle school. *The Journal of Early Adolescence*, 35(5–6), 790–816. <https://doi.org/10.1177/0272431614556890>
- Lane, A. K., Meaders, C. L., Shuman, J. K., Stetzer, M. R., Vinson, E. L., Couch, B. A., Smith, M. K., & Stains, M. (2021). Making a first impression: Exploring what instructors do and say on the first day of introductory STEM courses. *CBE—Life Sciences Education*, 20(1), ar7. <https://doi.org/10.1187/cbe.20-05-0098>
- Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., Fata-Hartley, C. L., Ebert-May, D., Jardeleza, S. E., & Cooper, M. M. (2016a). Characterizing college science assessments: The three-dimensional learning assessment protocol. *PLOS ONE*, 11(9). <https://doi.org/10.1371/journal.pone.0162333>
- Laverty, J. T. & Caballero, M. D. (2018b). Analysis of the most common concept inventories in physics: What are we assessing? *Physical Review Physics Education Research*, 14(1). <https://doi.org/10.1103/physrevphyseducre.14.010123>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- Lombardi, D., & Shipley, T. F. (2021). The curious construct of active learning. *Psychological Science in the Public Interest*, 22(1), 8–43. <https://doi.org/10.1177/1529100620973974>
- 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), 1–21. <https://doi.org/10.1186/s40594-015-0026-8>
- Matusovich, H. M., Paretto, M. C., McNair, L. D., & Hixson, C. (2014). Faculty motivation: A gateway to transforming engineering education. *Journal of Engineering Education*, 103(2), 302–330. <https://doi.org/10.1002/jee.20044>
- Matz, R. L. & Jardeleza, S. E. (2016b). Examining the role of leadership in an Undergraduate Biology Institutional Reform initiative. *CBE—Life Sciences Education*, 15(4). <https://doi.org/10.1187/cbe.15-10-0222>
- Matz, R. L., Fata-Hartley, C. L., Posey, L. A., Laverty, J. T., Underwood, S. M., Carmel, J. H., Herrington, D. G., Stowe, R. L., Caballero, M. D., Ebert-May, D., & Cooper, M. M. (2018a). Evaluating the extent of a large-scale transformation in Gateway Science Courses. *Science Advances*, 4(10). <https://doi.org/10.1126/sciadv.aau0554>
- McCormick, N. J., & Lucas, M. S. (2011). Exploring mathematics college readiness in the United States. *Current Issues in Education*, 14(1), 1. <https://cie.asu.edu/ojs/index.php/cieatasu/article/view/680>
- McCourt, J. S., Andrews, T. C., Knight, J. K., Merrill, J. E., Nehm, R. H., Pelletreau, K. N., Prevost, L. B., Smith, M. K., Urban-Lurain, M., & Lemons, P. P. (2017). What motivates biology instructors to engage and persist in teaching professional development? *CBE—Life Sciences Education*, 16(3), ar54. <https://doi.org/10.1187/cbe.16-08-0241>
- McNeill, K. L., Lowenhaupt, R., Cherbow, K., & Lowell, B. R. (2022). Professional development to support principals' vision of science instruction: Building from their prior experiences to support the science practices. *Journal of Research in Science Teaching*, 59(1), 3–29. <https://doi.org/10.1002/tea.21719>
- McPartlan, P., Thoman, D. B., Poe, J., Herrera, F. A., & Smith, J. L. (2022). Appealing to faculty gatekeepers: Motivational processes for intentions to adopt an evidence-based intervention. *BioScience*, 72(7), 664–672. <https://doi.org/10.1093/biosci/biac029>
- Miller, A. R., & Kastens, K. A. (2018). Investigating the impacts of targeted professional development around models and modeling on teachers' instructional practice and student learning. *Journal of Research in Science Teaching*, 55(5), 641–663. <https://doi.org/10.1002/tea.21434>
- National Research Council. (2012a). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. <https://doi.org/10.17226/13165>
- National Research Council. (2012b). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. The National Academies Press. [http://www.nap.edu/catalog.php?record\\_id=13362](http://www.nap.edu/catalog.php?record_id=13362)
- Nehm, R. H. (2019). Biology education research: Building integrative frameworks for teaching and learning about living systems. *Disciplinary and Interdisciplinary Science Education Research*, 1(1), 1–18. <https://doi.org/10.1186/s43031-019-0017-6>
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press. <http://www.nextgenscience.org/ngss-high-school-evidence-statements>
- Paprzycki, P., Tuttle, N., Czerniak, C. M., Molitor, S., Kadervaek, J., & Mendenhall, R. (2017). The impact of a Framework-aligned science professional development program on literacy and mathematics achievement of K-3 students. *Journal of Research in Science Teaching*, 54(9), 1174–1196. <https://doi.org/10.1002/tea.21400>
- Part, R., Perera, H. N., Marchand, G. C., & Bernacki, M. L. (2020). Revisiting the dimensionality of subjective task value: Towards clarification of competing perspectives. *Contemporary Educational Psychology*, 62, 101875. <https://doi.org/10.1016/j.cedpsych.2020.101875>
- Qu, J., Jiang, Q., Weng, F., & Hong, Z. (2007). A hierarchical clustering based on overlap similarity measure. In *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007)* (Vol. 3, pp. 905–910). IEEE.
- Ralph, V. R., Scharlott, L. J., Schafer, A. G. L., Deshaye, M. Y., Becker, N. M., & Stowe, R. L. (2022a). Advancing equity in STEM: The impact assessment design has on who succeeds in undergraduate introductory chemistry. *JACS Au*, 2(8), 1869–1880. <https://doi.org/10.1021/jacsau.2c00221>
- Ralph, V. R., Scharlott, L. J., Schwarz, C. E., Becker, N. M., & Stowe, R. L. (2022b). Beyond instructional practices: Characterizing learning environments that support students in explaining chemical phenomena. *Journal of Research in Science Teaching*, 59(5), 841–875. <https://doi.org/10.1002/tea.21746>
- Reinholz, D. L., & Apkarian, N. (2018). Four frames for systemic change in STEM departments. *International Journal of STEM Education*, 5(1), 1–10. <https://doi.org/10.1186/s40594-018-0103-x>
- Rodriguez, A. J. (2015). What about a dimension of engagement, equity, and diversity practices? A critique of the next generation science standards. *Journal of Research in Science Teaching*, 52(7), 1031–1051. <https://doi.org/10.1002/tea.21232>
- Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods*, 15(1), 85–109.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066x.55.1.68>
- Saldana, J. (2021). *The coding manual for qualitative researchers* (4th ed.). SAGE Publications Ltd.
- Sansom, R. L., Winters, D. M., & St. Clair, B. E., West, R. E., & Jensen, J. L. (2023). Factors that influence STEM faculty use of evidence-based instructional practices: An ecological model. *PLoS One*, 18(1), e0281290. <https://doi.org/10.1371/journal.pone.0281290>
- Schaeffer, M. S., & Levitt, E. E. (1956). Concerning Kendall's tau, a nonparametric correlation coefficient. *Psychological Bulletin*, 53(4), 338–346. <https://doi.org/10.1037/h0045013>
- Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Westview Press.
- Seymour, E., & Hunter, A.-B. (2019). *Talking about leaving revisited: Persistence, relocation, and loss in undergraduate STEM education*. Springer International Publishing.
- Shadle, S. E., Marker, A., & Earl, B. (2017). Faculty drivers and barriers: Laying the groundwork for undergraduate STEM education reform in academic departments. *International Journal of STEM Education*, 4(1), 1–13. <https://doi.org/10.1186/s40594-017-0062-7>
- Sosniak, L. A. (2006). Retrospective interviews in the study of expertise and expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 287–301). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816796.016>
- Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R. S., DeChenne-Peters, S. E., Eagan, M. K., Esson, J. M., Knight, J. K., Laski, F. A., Levis-Fitzgerald, M., Lee, C. J., Lo, S. M., McDonnell, L. M., McKay, T. A., Michelotti, N., Musgrove, A., Palmer, M. S., Plank, K. M., ... Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science*, 359(6383), 1468–1470. <https://doi.org/10.1126/science.aap8892>
- Stolk, J. D., Gross, M. D., & Zastavker, Y. V. (2021). Motivation, pedagogy, and gender: Examining the multifaceted and dynamic situational

- responses of women and men in college STEM courses. *International Journal of STEM Education*, 8(1), 1–19. <https://doi.org/10.1186/s40594-021-00283-2>
- Tasci, G. (2015). Criteria for high quality biology teaching: An analysis. *Educational Sciences: Theory & Practice*, 15(4), 4. <https://doi.org/10.12738/estp.2015.4.2420>
- The Carnegie Classification of Institutions of Higher Education. (n.d.). *Carnegie classifications*. Retrieved April 19, 2021, from <https://carnegieclassifications.iu.edu/index.php>
- 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>
- Wigfield, A., & Cambria, J. (2010). Expectancy-value theory: Retrospective and prospective. In T. C. Urdan & S. A. Karabenick (Eds.), *The decade ahead: Theoretical perspectives on motivation and achievement: Vol. 16 Part A* (pp. 35–70). Emerald Group Publishing Limited. [https://doi.org/10.1108/S0749-7423\(2010\)000016A005](https://doi.org/10.1108/S0749-7423(2010)000016A005)
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81. <https://doi.org/10.1006/ceps.1999.1015>
- Wigfield, A., Tonks, S., & Klauda, S. L. (2009). Expectancy-value theory. In K. R. Wenzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 55–75). Routledge/Taylor & Francis Group.
- Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2006). Development of achievement motivation. In *Handbook of child psychology: Social, emotional, and personality development, Vol. 3, 6th ed* (pp. 933–1002). Wiley.
- Wilson-Kennedy, Z. S., Huang, L., Kennedy, E., Tang, G., Kanipes, M. I., & Byrd, G. S. (2019). Faculty motivation for scholarly teaching and innovative classroom practice—An empirical study. In M. Schultz, S. Schmid, & G. A. Lawrie (Eds.), *Research and practice in chemistry education: Advances from the 25th IUPAC International Conference on Chemistry Education 2018* (pp. 65–88). Springer. [https://doi.org/10.1007/978-981-13-6998-8\\_5](https://doi.org/10.1007/978-981-13-6998-8_5)
- Yang, Y., Liu, X., & Gardella, J. A., Jr. (2020). Effects of a professional development program on science teacher knowledge and practice, and student understanding of interdisciplinary science concepts. *Journal of Research in Science Teaching*, 57(7), 1028–1057. <https://doi.org/10.1002/tea.21620>
- Yoho, R., Foster, T., Urban-Lurain, M., Merrill, J., & Haudek, K. C. (2019). Interdisciplinary insights from instructor interviews reconciling “structure and function” in biology, biochemistry, and chemistry through the context of enzyme binding. *Disciplinary and Interdisciplinary Science Education Research*, 1(1), 1–17. <https://doi.org/10.1186/s43031-019-0016-7>
- Zagallo, P., McCourt, J., Idsardi, R., Smith, M. K., Urban-Lurain, M., Andrews, T. C., Haudek, K., Knight, J. K., Merrill, J., Nehm, R., Prevost, L. B., & Lemons, P. P. (2019). Through the eyes of faculty: Using personas as a tool for learner-centered professional development. *CBE—Life Sciences Education*, 18(4), ar62. <https://doi.org/10.1187/cbe.19-06-0114>

## Publisher's Note

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

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

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

---

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

---