

# Crossing Boundaries: Steps Toward Measuring Undergraduates' Interdisciplinary Science Understanding

Brie Tripp, Sophia A. Voronoff, and Erin E. Shortlidge\*

Department of Biology, Portland State University, Portland, OR 97201

## ABSTRACT

A desired outcome of education reform efforts is for undergraduates to effectively integrate knowledge across disciplines in order to evaluate and address real-world issues. Yet there are few assessments designed to measure if and how students think interdisciplinarily. Here, a sample of science faculty were surveyed to understand how they currently assess students' interdisciplinary science understanding. Results indicate that individual writing-intensive activities are the most frequently used assessment type (69%). To understand how writing assignments can accurately assess students' ability to think interdisciplinarily, we used a preexisting rubric, designed to measure social science students' interdisciplinary understanding, to assess writing assignments from 71 undergraduate science students. Semistructured interviews were conducted with 25 of those students to explore similarities and differences between assignment scores and verbal understanding of interdisciplinary science. Results suggest that certain constructs of the instrument did not fully capture this competency for our population, but instead, an interdisciplinary framework may be a better model to guide assessment development of interdisciplinary science. These data suggest that a new instrument designed through the lens of this model could more accurately characterize interdisciplinary science understanding for undergraduate students.

## INTRODUCTION

The interplay of science, technology, engineering, and mathematics (STEM) fields has impacted research in profound ways, with new discoveries accelerating scientific advancements (President's Council of Advisors on Science and Technology [PCAST], 2012). The integration of STEM with disciplines such as economics and sociology has created new interdisciplinary (ID) fields that hold tremendous promise for surmounting societies' most vexing challenges (National Research Council [NRC], 2009). Also driving these important avenues of study are transformations in how scientists communicate and collaborate across disciplines (American Association for the Advancement of Science [AAAS], 2011). Novel theories and methods often arise from these interactions, warranting continued ID efforts to advance scientific fields and address complex issues (NRC, 2003, 2009; AAAS, 2011). As such, future scientists must be equipped with a skill set that enables them to effectively address problems that span multiple disciplinary domains. However, undergraduate education has not entirely kept up with this need, as universities have been slower in engaging students in ID practices (NRC, 2003, 2009; AAAS, 2011). Accordingly, national calls have developed mandates for improving undergraduate education to match ID scientific advancements (NRC, 2003, 2009; AAAS, 2011; PCAST, 2012). The NRC (2003) specifically highlights this need for life science majors: "Connections between biology and other scientific disciplines need to be developed and reinforced so that interdisciplinary thinking and work become second nature" (p. 1).

In response to calls for reform, the AAAS (2011) outlined several core competencies in the meeting report *Vision and Change in Undergraduate Biology Education* to

**Tammy Long**, *Monitoring Editor*

Submitted Sep 4, 2019; Revised Dec 11, 2019;  
Accepted Jan 2, 2020

CBE Life Sci Educ March 1, 2020 19:ar8

DOI:10.1187/cbe.19-09-0168

\*Address correspondence to: Erin E. Shortlidge  
(eshortlidge@pdx.edu).

© 2020 B. Tripp *et al.* CBE—Life Sciences Education © 2020 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

better prepare undergraduate biology students for the increasingly ID workforce. The ability to “tap into the interdisciplinary nature of science” is one of these competencies that science educators have been working to incorporate into curricula; however, this competency can be difficult to operationalize and evaluate (Tripp and Shortlidge, 2019). If science educators are tasked with instilling this proficiency, we must find ways to assess whether students are meeting this benchmark.

### Assessment of Interdisciplinary Science Understanding

Instructors are at the forefront of designing and teaching curricula that meet ID learning outcomes with the expectation that they will assess whether students are meeting these goals. Therefore, shedding light on instructor assessment practices can move us closer toward meeting the ID recommendations outlined in *Vision and Change* (AAAS, 2011). A small number of studies have been published on efforts to measure ID science competencies through concept maps and writing activities. Borrego *et al.* (2009) tested a rubric (Besterfield-Sacre *et al.*, 2004) to assess engineering students’ ability to integrate ID knowledge through concept maps. They had students schematically represent their knowledge of integration by creating a hierarchy of concepts across disciplines, associating subgroups branching off each concept, and pairing these ideas with cross-linked arrows to represent relationships. They found that the assessment tool did not produce accurate or reliable results in scoring students’ ID knowledge based on variability in interpretation of students’ work. Although concept maps are useful in particular environments, they can fall short when asking students to exhibit a deeper understanding of *why* conceptual knowledge is connected across seemingly disparate disciplinary fields (Balgopal *et al.*, 2012).

Several studies have developed assessment tools to score writing activities in specific environments, but these instruments were either targeted for one particular course without additional validation studies from separate populations (Balgopal *et al.*, 2012, 2017) or focused on integrated learning within one discipline (Besterfield-Sacre *et al.*, 2004; Chan *et al.*, 2010). However, writing activities may be a plausible assessment strategy when tasking students to connect similarities and differences in jargon, methods/methodologies, and concepts and ideas across multiple disciplines (Boix Mansilla *et al.*, 2009). Writing can promote reflection and encourage students to be critical of their own understanding while allowing space for affective learning to enhance greater literacy on real-world issues (Connolly and Vilardi, 1989; Rivard 1994; Keys, 1999; Balgopal *et al.*, 2012). For example, a pedagogy known as “writing-to-learn” was adapted to extend students’ learning beyond rote memorization and simplified connections (Connolly, 1989; Rivard, 1994). Writing-to-learn, a constructivist teaching strategy that allows students to construct their own understanding by first thinking and writing about a topic before actually engaging in content-related activities, has been more recently adopted in the sciences (Carlson, 2007; Balgopal and Wallace, 2009; Balgopal *et al.*, 2012, 2017). Writing-to-learn aligns with work suggesting that students will likely need to think interdisciplinarily before engaging in ID science research (Tripp and Shortlidge, 2019). Yet it remains relatively unknown whether instructors are encouraging ID science thinking, and if so, how they assess this ability in an undergraduate classroom.

To address this, the first part of this study examined how instructors assess ID science competencies. We then used these results to guide the development of a writing activity to be scored with an ID rubric.

### A Theoretical Model and an Interdisciplinary Rubric

Our previous work outlined a model, the Interdisciplinary Science Framework (IDSF), to guide instructors on factors to consider when developing ID curricula and assessing student understanding of ID science (Tripp and Shortlidge, 2019). As a step in building this model, we surveyed faculty who teach science courses regarding how they define ID science ( $n = 184$ ). By synthesizing these definitions and studying the ID literature, we established five main categories that comprise ID science understanding: 1) disciplinary humility, 2) disciplinary grounding, 3) different research methods, 4) advancement through integration, and 5) collaboration.

The IDSF categories disciplinary grounding and integration were derived from criteria theorized as pivotal for interdisciplinarity in the social sciences (Boix Mansilla and Duraisingh, 2007). These researchers developed a rubric to score social science and humanities students’ understanding of the constructs, along with two additional constructs, “purposefulness” and “critical awareness” (Boix Mansilla *et al.*, 2009). One study examined the rubric’s functionality on grant proposals submitted to the National Science Foundation’s former Interdisciplinary Graduate Engineering Research Traineeship (IGERT) program (Borrego and Newswander, 2010). Researchers used the rubric to compare grants submitted in response to the IGERT solicitation to identify learning outcomes for proposed ID graduate programs. The researchers’ findings suggested that the constructs of the rubric, although applicable to the physical sciences, needed amendment to fully capture learning outcomes for graduate students in STEM fields (Borrego and Newswander, 2010). Here, we expand this work by testing the same rubric’s (Boix Mansilla *et al.*, 2009) ability to measure responses to a situated undergraduate writing assignment.

Although the rubric was developed from faculty feedback across many STEM and non-STEM disciplines, the designers limited validation of the data to students in the social sciences and humanities (Boix Mansilla *et al.*, 2009). Instead of initially modifying this rubric to align with the ID science-focused IDSF, we chose to maintain the integrity of the instrument by using it as published. Had we changed the criteria in the rubric without first testing the validity of the data collected, our findings would potentially be invalidated (Stangor, 2014). We hypothesized that the results from testing the rubric would not only assist in understanding how students conceptualize ID science, but also test for evidence of validity for the IDSF model.

### Research Aims

In this study, we first aimed to reveal how instructors currently assess ID science understanding, and we use this information to inform the development of an activity to measure undergraduate ID science understanding. We then scored this activity with a previously developed ID rubric to test whether the rubric produced valid data in our population. We examined whether the rubric fully captured students’ ID understanding by conducting interviews to holistically probe how students perceive ID science. Finally, we used these interviews to test for evidence

of validity for the theoretically driven IDSF. Specifically, we asked the following research questions:

1. How do instructors typically assess undergraduate students' conceptualization of interdisciplinary science?
2. In what ways can a previously developed rubric measure undergraduate students' interdisciplinary science understanding?
  - a. Which aspects of the rubric are more or less difficult for students to communicate, and does this vary by course?
  - b. Can the rubric accurately measure undergraduate students' interdisciplinary science understanding?
3. How do undergraduate students perceive interdisciplinary science?

## METHODS

### Research Question 1. How Do Instructors Typically Assess Undergraduate Students' Conceptualization of Interdisciplinary Science?

**Survey Recruitment.** To gauge how science instructors assess ID science understanding, we conducted a Web-based search for participants that spanned STEM departments across the United States. We compiled an email list of potential participants and sent individual and Listserv emails requesting anonymous participation in a Qualtrics survey regarding ID science. Individuals were invited to participate if they 1) held a faculty position at an academic institution and 2) had a position located in a science department. The survey items underwent iterative revision based on feedback from multiple researchers (including authors B.T., S.A.V., and E.E.S.).

This portion of the study was conducted under exempt status at Portland State University (IRB no. 174219).

**Data Collection.** The survey asked participants a series of demographic questions, one binary (yes or no) question: "Do you teach courses that you consider interdisciplinary?"; and two open-ended questions: "How do you define interdisciplinary science?" (results can be found in Tripp and Shortlidge, 2019), and "Please explain how you assess learning outcomes related to students' understanding of interdisciplinary science" (presented in this study). We used inductive content analysis (Patton, 1990) to evaluate responses to the latter open-ended survey question. Two researchers (including B.T.) compiled responses into a list, which was subsequently organized into categories of similar assessment strategies. Because the survey responses were often provided in an itemized format containing the same or similar words in each response (e.g., Survey Participant 1: quizzes, tests, oral presentations; Survey Participant 2: quizzes, exams, essays, verbal presentations), we categorized associated words into a code resulting in multiple codes with related words per code. Thus, very little interpretation was used in the development of the code list due to the categorizing of exact or associated words. All survey responses were coded to consensus. We then condensed interrelated codes into overarching themes that holistically represented the codes. A researcher uninvolved in the initial coding process (E.E.S.) independently evaluated 20% of the data at random—checking for accuracy and appropriateness of the coding scheme—as an additional measure to support the validity of our analysis.

**Essay Assignment Development.** On the basis of results from our faculty survey, we created a writing assignment to collect a sample of science students' abilities to think interdisciplinarily. Additional reasons for developing a writing assignment were threefold: writing activities are being adopted at a higher rate in science to encourage critical-thinking skills (Connolly and Vilardi, 1989; Carlson, 2007; Balgopal and Wallace, 2009; Balgopal *et al.*, 2012, 2017); we sought to engage students in connecting multiple disciplines in a cohesive manner by first thinking through an ID lens (Tripp and Shortlidge, 2019); and the assignment served as an artifact to measure students' ID understanding using the aforementioned rubric developed by Boix Mansilla *et al.* (2009).

One way to possibly enhance student ability to meaningfully connect disciplines is through real-world applications of ID science, as ID work is how we truly solve complicated issues in society (Tripp and Shortlidge, 2019). Therefore, in this study, we developed essay prompts that task students to ponder real-world problems that inherently require multiple disciplines to address. Two authors (B.T. and E.E.S.) iteratively developed and revised essay prompts in collaboration with the instructor(s) of each course to ensure that content aligned with the course subject matter. Although the context of the prompt varied between courses, student instructions for completing the assignment remained consistent across courses, and all instructors incorporated the assignment into the grading scheme of the course (see Supplemental Material 1 for example prompts). We intentionally worded the prompts to encourage students to meet each construct in the rubric (rubric details are discussed later). The research team collaboratively discussed different types of student knowledge that could potentially satisfy understanding of the rubric's constructs based on the information given in the prompt.

### Research Question 2. In What Ways Can a Previously Developed Rubric Measure Undergraduate Students' Interdisciplinary Science Understanding?

#### 2a. Which Aspects of the Rubric Are More or Less Difficult for Students to Communicate, and Does This Vary by Course?

**Data Collection: Recruitment.** We recruited undergraduate students from four upper-division natural and physical science courses at a large northwestern public university in 2017–2018. We targeted students in upper-division courses because ID understanding is partially contingent upon higher-order thinking, and students at the beginning of their academic careers may not have had the experience or time to fully develop these skills (Tripp and Shortlidge, 2019).

The essay assignment was given to all students enrolled in each course, and scores were incorporated into their overall course grades. One week before the assignment, we made a class announcement requesting consent from students to use their essay responses for education research; written consent forms were disseminated and collected in class. The majority of students consented to their responses being included in the study ( $n = 71$ , 99% consent rate). Although the assignment was part of the final grade, student involvement in this study was completely voluntary and participation remained anonymous to the instructor. Students were given 1 week to complete the individual assignment and were allowed to use any resources they chose to complete the essay. The assignment was worth 10–15% of final course grades.

We recruited students for interviews during the same class announcement and provided a sign-up sheet for students interested in participating. A follow-up email was sent to those who volunteered to participate in interviews. There were no students with dual enrollment in any of the courses.

This portion of the study was conducted under exempt status at Portland State University (IRB no. 163998).

*The Rubric.* The rubric was designed to reconcile “rhetorical, theoretical, and methodological” differences among disciplines, with an aim of discerning student competencies in ID understanding (Boix Mansilla *et al.*, 2009). It was intended to be an adaptable assessment tool to guide instructors in qualities of students’ understanding of interdisciplinarity based on four constructs: purposefulness, disciplinary grounding, integration, and critical awareness (Table 1).

We performed a pilot test of the assignment on a group of students from the same population as our study, and subsequently scored essay responses with the rubric ( $n = 13$ ). The rubric was withheld from the students based on its intended design for practitioners’ use. We also were not able to discern whether the authors of the rubric had provided students with a version of the rubric in their study. Based on verbal feedback and analysis of scores on assignments, it was evident that students needed more guidance in writing an essay that demonstrated how they conceptualize ID connections. The full rubric intended for instructors had guiding questions within the document (see Boix Mansilla *et al.*, 2009, for full rubric). We decided to include these guiding questions alongside the essay prompts for our subsequent research to help students in meeting expectations (Table 1).

Next, we aimed to establish evidence of validity of data collected in our population by applying the rubric to the essay assignment in the four courses described in this study (Barbera and VandenPlas, 2011; American Educational Research Association, American Psychological Association, and National Council on Measurement in Education [AERA *et al.*], 2013).

The conceptual foundations and core elements of the rubric were strictly followed to ensure fidelity of implementation, with minimal adaptations to the four constructs and scoring metrics. However, there were two criteria—3.2: Is there an integrative device or strategy (i.e., metaphor, or analogy)? and 4.2: Does the paper exhibit self-reflection (e.g., metacognition)?—that did not fit the context of this study based on the responses from students in the pilot test. We tasked students with writing an essay to governmental bodies or scientific enterprises, thus, usage of analogies and metaphors in criterion 3.2 would not be appropriate for this population. Criterion 4.2 was extremely similar to criterion 4.1, and we had a difficult time disaggregating their meaning, as did students in the pilot study. Therefore, criterion 3.2 was excluded from scoring the essays and 4.2 was merged with 4.1 (Table 1).

*Rubric Scoring.* Each construct (purposefulness, disciplinary grounding, integration, and critical awareness) was scored on a four-point scale as outlined by the designers of the rubric: naïve (1), novice (2), apprentice (3), and mastery (4). There were also several detailed criteria associated with the four constructs to aid instructors in assessing students’ understanding of interdisciplinarity (Table 1). However, it was unclear whether these criteria were to be scored on the 1–4 scale individually or whether the scoring metric was exclusively reserved for the construct as a whole. Thus, we decided to score each criterion on the 1–4 scale and then calculate the average of all criteria within one construct, resulting in one score per construct. An average of criteria scores was necessary, because there were different numbers of criteria under each construct, and we wanted to ensure they were weighted the same (e.g., having two criteria in critical awareness does not make that construct less important than the integration construct containing four criteria). There were some criteria that the coding researchers were not able to assess on the 1–4 whole-number scale; thus we resorted to using a rational (decimal) number scores (i.e., 1.5, 2.5, 3.5). For example, a student could have received a subscore of 1.5 due to

**TABLE 1. Shortened rubric provided to students<sup>a</sup>**

Rubric elements	Criteria	Guiding questions
Purposefulness	1.1	Is there a clearly stated purpose that calls for an integrative approach and a clear rationale or justification for taking this approach?
	1.2	Does the paper use the writing genre effectively to communicate with its intended audience?
Disciplinary grounding	2.1	Does the paper use disciplinary knowledge accurately and effectively (e.g., concepts, perspectives, findings, examples, relevant and credible sources)?
	2.2	Does the paper use disciplinary methods accurately and effectively (e.g., experimental design)?
Integration	3.1	Does the paper include selected disciplinary perspectives and insights from two or more disciplinary traditions presented in the course or from elsewhere that are relevant to the paper’s purpose?
	3.2 <sup>b</sup>	Is there an integrative device or strategy (i.e., metaphor or analogy)?
	3.3	Is there a sense of balance in the overall composition of the piece with regard to how disciplinary perspectives are brought together to advance the purpose of the piece?
	3.4	Do the conclusions drawn by the paper indicate that understanding has been advanced by the integration of disciplinary views (e.g., the paper takes full advantage of the opportunities presented by the integration of disciplinary insights to advance its intended purpose both effectively and efficiently; integration may result in novel or unexpected insights)?
Critical awareness	4.1 <sup>c</sup>	Does the paper exhibit awareness of the limitations and benefits of the contributing disciplines?
	4.2 <sup>c</sup>	Does the paper exhibit self-reflection (e.g., metacognition)?

<sup>a</sup>Adapted from Boix Mansilla *et al.* (2009).

<sup>b</sup>Excluded from scoring.

<sup>c</sup>Merged.

the fact that they were exhibiting knowledge in between a naïve (1) and novice (2) understanding. After each construct score was calculated, we added these numbers together for a total score out of 16 possible points for the assignment.

Two researchers (B.T. and S.A.V.) randomly selected and scored a subset of student essays from each of the four courses based on this scoring method until there was consistency in the scoring of students' work. A second subset of essays were independently scored by both researchers, who then reconvened and discussed each construct and associated criteria until consensus was reached. We continued this iterative process with 63% of all essays. A final subset of essays was scored and Cohen's kappa coefficient was obtained using R Studio ( $\kappa = 0.77$ ; R Studio Team, 2019). B.T. independently scored the remaining essays, occasionally having S.A.V. check for accuracy in scores from essays that were difficult to score.

*Statistical Analysis.* Statistical analyses were performed to explore differences in student performance on the essay based on the rubric constructs and the courses in which students were enrolled. One-way analyses of variance (ANOVAs) were used to identify any statistically significant differences in student performance based on construct and total essay scores. Welch's one-way test for unequal variance was used to account for a significant Levene's test for overall mean construct scores. A one-way ANOVA was also used to detect statistically significant differences among student performances based on each construct by course. Welch's one-way test for unequal variance was used on the integration and critical awareness constructs due to a significant Levene's test. Tukey's HSD post hoc analyses were conducted on each statistical measurement to further identify significance between groups ( $p < 0.05$ ). Effect sizes were calculated with eta-squared ( $\eta^2$ ) and interpreted according to Maher *et al.* (2013): small effect = 0.01, medium effect = 0.06, large effect = 0.14. All statistical tests were performed in R Studio (R Studio Team, 2019).

### **Research Question 2b. Can the Rubric Accurately Measure Undergraduate Students' Interdisciplinary Science Understanding?**

*Interviews.* To explore the breadth of students' understanding of ID science, a researcher (B.T.) conducted semistructured interviews. The interview questions were formulated to organically investigate student understanding of interdisciplinarity unrelated to the rubric or essay assignment (see Supplemental Material 2 for interview questions). We asked participants about their general perceptions of the course in which the required writing assignment was administered. We also inquired about students' experiences with research, how they viewed scientific disciplines and ID science, and the value they placed on both. The questions were first piloted on a group of eight education researchers (ranging from undergraduates to faculty) to assess the quality, accuracy, and intent of each question. After thorough discussion and deliberation, 20 questions were selected for this study. All questions remained consistent across the semistructured interviews. Each interview was recorded and transcribed verbatim, with interviews lasting an average of 30 minutes. For simplicity in reporting and protection of participant identities, all interviewees were given gender-neutral pseudonyms and are reported using nonbinary pronouns.

*Evidence of Convergent Validity through Matched Data.* When initially reading through the interviews, we noticed similarities and differences between student essay responses and how they articulated ID science in their interviews. As there was a subset of students who participated in both the writing assignment and an interview, we were able to better understand how these students were interpreting disciplinary grounding, integration, and critical awareness constructs and the associated criteria from the rubric. Thus, we examined the data for evidence of convergent validity. Convergent evidence is one type of validity that evaluates relationships between test scores and other external variables to assess the same or similar constructs (AERA *et al.*, 2013). We hypothesized that the rubric constructs should be related to how students articulate ID science understanding in their interviews, if the constructs were operating as the designers intended.

As interview questions were not specifically designed to address the rubric constructs, we did not have comparable matched data for the purposefulness construct in the interview responses. The criteria for purposefulness were very specific to essays—framing the problem and using a writing genre to communicate to an appropriate audience—and would likely not add to our understanding of how students conceptualize ID science. Thus, we omitted purposefulness from our matched writing assignment and interview data analysis.

*Scoring Interviews with the Rubric.* We identified responses in the interviews that aligned with each of the constructs in the rubric, ultimately scoring the interviews binarily (e.g., “yes” if the student exhibited the rubric construct disciplinary grounding in the interview [unprompted] or “no” if the student did not). We hypothesized that students who scored high on a particular rubric construct would also communicate an advanced level of understanding in their interviews regarding that construct. Likewise, we expected that low essay scores on a construct would be mirrored by little or no expression of the concept in the interviews. This method uniquely allowed us to examine evidence of convergent validity of data from two different measurements designed to assess the same concept of ID science understanding.

### **Research Question 3. How Do Undergraduate Students Perceive Interdisciplinary Science?**

To examine student perceptions of ID science that fell outside of the rubric constructs, we reanalyzed the interview data both inductively and deductively.

*Inductive Analysis.* Using holistic coding (Saldaña, 2015)—a method that applies a single code to each large unit of data to capture an overall sense of emergent content—three researchers (including B.T. and S.A.V.) performed inductive content analysis (Patton, 1990)—an analysis that uses the data to derive the structure of investigation—by systematically listing all emergent categories from 30% of the interviews. We reorganized and condensed similar categories into general codes. Once the final code list was complete, researchers independently coded a new subset of interviews with the codebook (20%) and reconvened to discuss new codes and reach consensus on coding interpretation. This process of iteratively coding and revising the codebook was repeated until we reached data saturation (Fusch and Ness, 2015). Next, a new subset of the

remaining interviews (20%) were independently coded and analyzed with a Fleiss's kappa coefficient  $>0.60$  ( $\kappa = 0.63$ ) using R Studio (R Studio Team, 2019). The remaining interviews were coded to consensus.

**Deductive Analysis.** After we completed the matched data and inductive interview analyses for this study, our work on the IDSF came to completion and was published (Tripp and Shortlidge, 2019). This provided an opportunity to learn more from the student interviews and test the robustness of the IDSF by recoding the interviews based on the five pillars of ID science understanding: disciplinary grounding, different research methods, integration, collaboration, and disciplinary humility (Tripp and Shortlidge, 2019). As the IDSF was partially developed through faculty perspectives of ID science, we could test for evidence of convergent validity of data for the model through student perspectives of this competency. Three researchers (including B.T. and S.A.V) performed deductive content analysis (Patton, 1990)—a method that tests existing categories or theories in a novel context—by reviewing all student interviews and applying codes to the responses that aligned with the five criteria in the IDSF. All interviews were coded to consensus, and coding analyses were conducted in MAXQDA (VERBI software, Berlin, Germany).

## RESULTS

### Research Question 1. How Do Instructors Typically Assess Undergraduate Students' Conceptualization of Interdisciplinary Science?

**Survey.** From the survey recruitment effort, 186 individual faculty members completed all survey questions. We excluded responses that were incomplete or written in a way that indicated participants did not understand the question as intended. In response to the question, "Do you teach courses that you consider interdisciplinary?," 45% ( $n = 84$ ) selected "yes." Of these 84 participants, 81% ( $n = 68$ ) also responded to the follow-up question, "Please explain how you assess these learning outcomes related to students' understanding of the interdisciplinary nature of science" (see Supplemental Material 3 for demographics and Supplemental Material 4 for survey questions).

The top three themes reported by faculty were coded as Writing Activities (e.g., essays, journal reflections; 69%), Traditional (e.g., quizzes, exams, and homework that were not described by the survey respondent as completed individually or in a group; 34%), and Group Work (e.g., group presentations, group projects; 34%; Table 2). Many faculty members listed assessment strategies that fell into more than one theme, hence the percentages sum to greater than 100%.

### Research Question 2. In What Ways Can a Previously Developed Rubric Measure Undergraduate Students' Interdisciplinary Science Understanding?

#### 2a. Which Aspects of Rubric Are More or Less Difficult for Students to Communicate, and Does This Vary by Course?

**Student Performance Based on Rubric Construct Scores.** To evaluate which constructs in the rubric were more or less difficult for students to communicate in their essay responses, we compared student performance (by construct) using the 1–4 scoring metric ( $n = 71$ ; Table 3; Figure 1). There was a significant difference between students' scores by construct ( $F(3, 280) =$

**TABLE 2. Coding rubric for survey question "Please explain how you assess learning outcomes related to students' understanding of interdisciplinary science" ( $n = 68$ )**

Themes	Examples	Participants % ( $n = 68$ ) <sup>a</sup>
1. Writing activities		69 (47)
a. Writing assignments	Essays/papers	51 (35)
b. Self-reflection	Journals	6 (4)
	Reflection assignments	
2. Traditional	Unspecified as individual or group	34 (23)
	Exams	
	Quizzes	
	Homework assignments	
3. Group work	Two or more students	34 (23)
	Communication/discussion	
	Group research/projects	
	Problem-based learning	
	Group presentation	

<sup>a</sup>Percentages are greater than 100% due to responses being coded into multiple themes.

$6.149, p = 0.00057, \eta^2 = 0.062$ ). Pairwise comparisons revealed that, overall, students scored significantly higher in the purposefulness construct than on integration or critical awareness ( $p = 0.0025$  and  $p = 0.0139$ , respectively), and on average, scored significantly higher on the disciplinary grounding construct than on integration ( $p = 0.0185$ ).

**Student Performance Based on Course.** To disaggregate differences between student ID science understanding by course, we first compared total average essay scores of students from each course (Figure 2). There was a significant difference between student scores by course ( $F(3, 67) = 3.69, p = 0.016, \eta^2 = 0.142$ ). Pairwise comparisons indicated that mean essay scores of students in Chemical Ecology were significantly higher than those of students in Environmental Restoration ( $p = 0.0187$ ), with no significant differences between other courses.

**Student Performance Based on Construct by Course.** Next, we analyzed average student performance on each rubric construct by course, illustrating an overall significant difference between courses in disciplinary grounding ( $F(3, 68) = 14.5, p < 0.0001, \eta^2 = 0.329$ ), integration ( $F(3, 68) = 19.2, p < 0.0001, \eta^2 = 0.401$ ), and critical awareness ( $F(3, 68) = 8.38, p = 0.0003, \eta^2 = 0.187$ ), with no significant differences between courses in purposefulness (Figure 3A). For disciplinary grounding, pairwise comparisons revealed students enrolled in Chemical Ecology ( $p < 0.0001$ ), Biochemical Virology ( $p = 0.002$ ), and Plant Systematics ( $p = 0.044$ ) scored significantly higher than students in Environmental Restoration, with no significant differences between the former three courses (Figure 3B). For integration, students enrolled in Biochemical Virology and Chemical Ecology significantly outperformed students in Plant Systematics ( $p = 0.0207$  and  $p = 0.0138$ , respectively), as well as those enrolled in Environmental Restoration ( $p < 0.0001$  for both courses; Figure 3C). For critical awareness, students enrolled in Chemical Ecology and Environmental Restoration scored significantly higher than those enrolled in Plant Systematics ( $p = 0.006$  and  $p = 0.016$ ,

TABLE 3. Course characterization of four upper-division natural and physical science courses

Course	Format	Credits	Total no. of essay participants	Total no. of interview participants	Disciplinary or ID; Course-listed departments	Instructors
Biochemical Virology	Lecture	1	11	4	ID; Biology and Chemistry	1 Biochemist 1 Biologist
Chemical Ecology	Lecture + research-based lab	3	13	8	ID; Biology and Chemistry	1 Chemist 1 Biologist
Environmental Restoration	Lecture	3	32	6	ID; Environmental Sciences and Management	1 Ecologist
Plant Systematics	Lecture + traditional lab	4	15	7	Disciplinary; Biology	1 Biologist

respectively; Figure 3D). Overall, students enrolled in the Chemical Ecology course scored significantly higher in every construct (except purposefulness) compared with at least one other course.

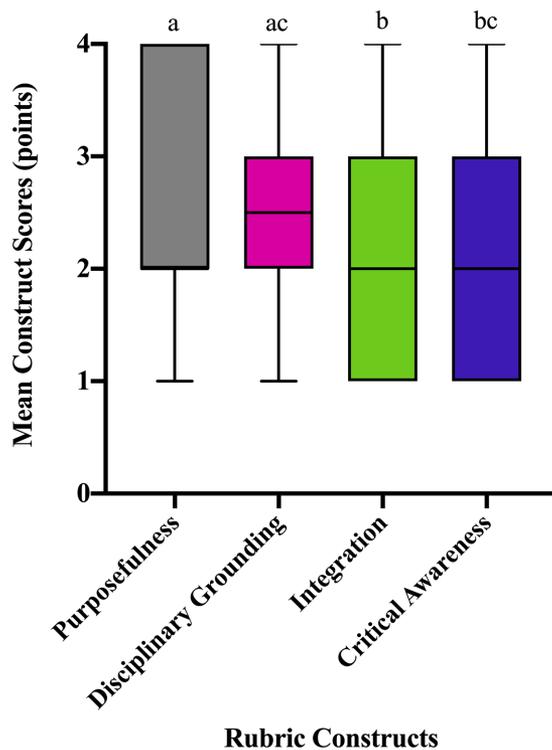


FIGURE 1. Box plots compare student overall mean construct scores ( $n = 71$ ). Nonidentical letters above bars represent significant ( $p < 0.05$ ) differences among construct scores (as determined by ANOVA and post hoc pairwise comparisons using Tukey's HSD). A one-way Welch's ANOVA detected a significant difference between mean construct scores ( $F(3, 280) = 6.149, p = 0.00057, \eta^2 = 0.062$ ). Tukey's post hoc analyses reveal that students scored significantly higher on purposefulness than integration and critical awareness ( $p = 0.0025$  and  $p = 0.0139$ , respectively), with no significant differences between the latter two constructs. Students performed significantly better on disciplinary grounding than integration ( $p = 0.0185$ ), with no significant differences between disciplinary grounding and purposefulness. Box: 25th to 75th percentile; bars: minimum and maximum values. The error bars represent the standard error of the mean.

*Research Question 2b. Can the Rubric Accurately Measure Undergraduate Students' Interdisciplinary Science Understanding?*

*Interviews.* We interviewed a subset of students from each course who had completed the essay assignment to test whether the rubric accurately and adequately captured ID science understanding in our population. In total, 25 of the 71 students

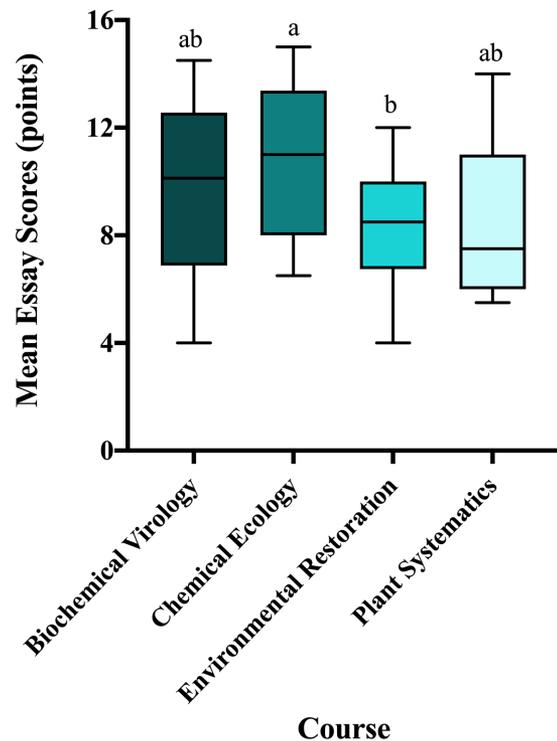
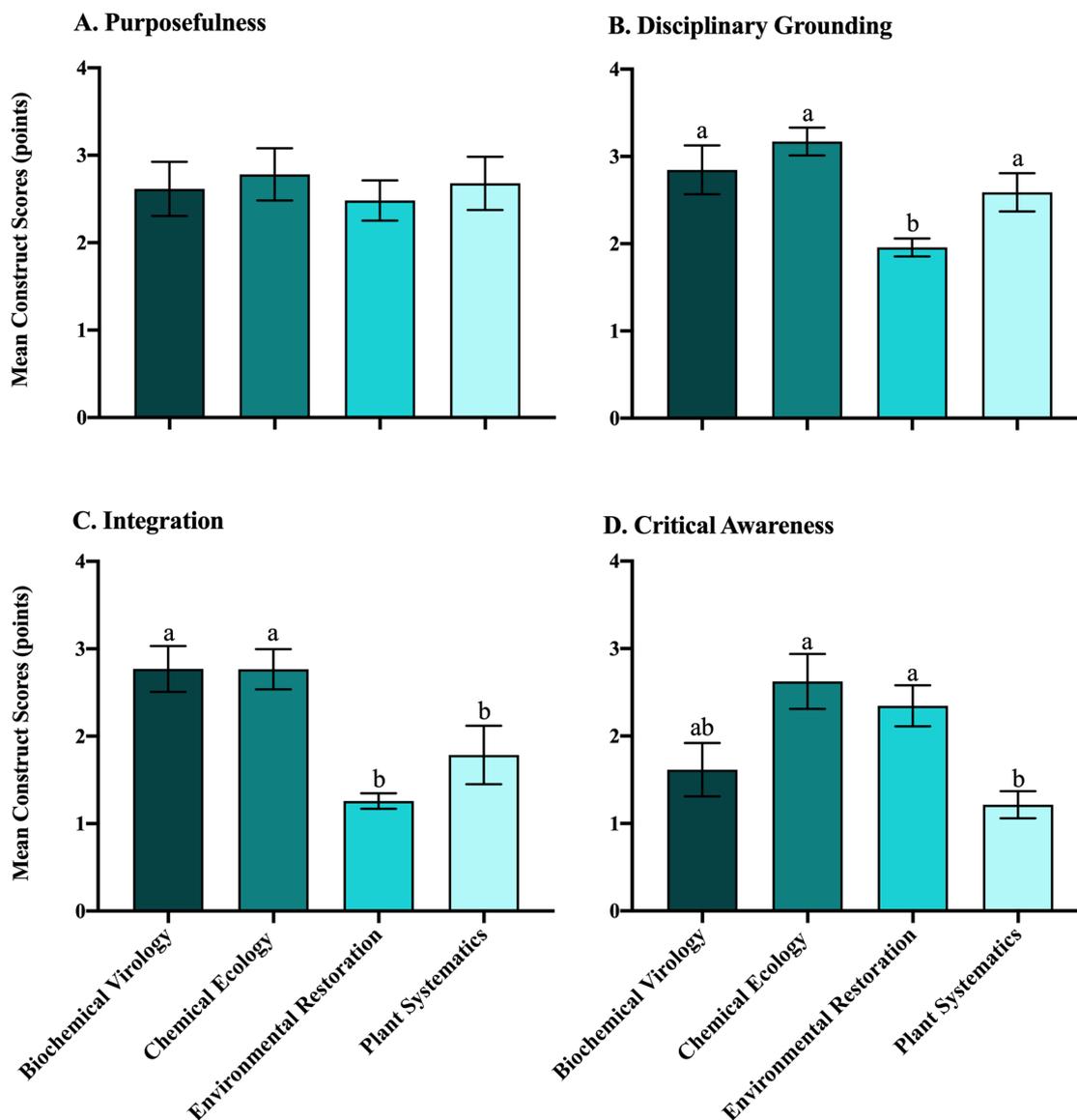


FIGURE 2. Box plots compare students' mean essay scores across four upper-division courses ( $n = 71$ ). Nonidentical letters above bars represent significant ( $p < 0.05$ ) differences among courses (as determined by ANOVA and post hoc pairwise comparisons using Tukey's HSD). One-way ANOVA revealed a significant difference between mean construct scores ( $F(3, 67) = 3.691, p = 0.016, \eta^2 = 0.142$ ). A Tukey's post hoc test indicated a significant difference in mean essay scores between Chemical Ecology and Environmental Ecology ( $p = 0.0187$ ), with no significant differences between other courses. Box: 25th to 75th percentile; bars: minimum and maximum values.

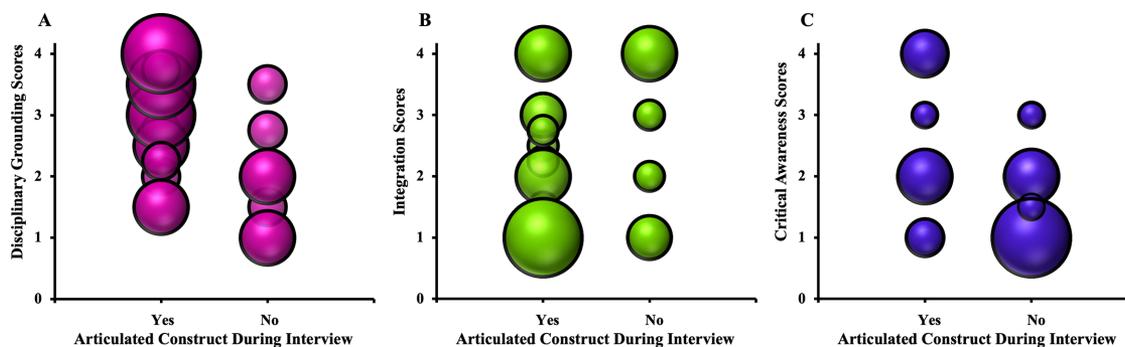


**FIGURE 3.** Comparison of mean construct scores for students enrolled in four courses ( $n = 71$ ). Nonidentical letters above bars represent significant ( $p < 0.05$ ) differences among courses within each construct (as determined by ANOVA and post hoc pairwise comparisons using Tukey's HSD). One-way ANOVA indicated a significant difference between course scores based on the constructs disciplinary grounding ( $F(3, 68) = 14.5, p < 0.0001, \eta^2 = 0.329$ ), integration ( $F(3, 68) = 19.2, p < 0.0001, \eta^2 = 0.401$ ), and critical awareness ( $F(3, 68) = 8.38, p = 0.0003, \eta^2 = 0.187$ ; Welch's ANOVA for unequal variances reported based on significant Levene's test for integration and critical awareness). Tukey's post hoc tests: (A) construct purposefulness: no significant differences in student scores across courses; (B) construct disciplinary grounding: students in Chemical Ecology, Biochemical Virology, and Plant Systematics score significantly higher than students in Environmental Restoration ( $p < 0.0001, p = 0.0024, \text{ and } p = 0.0435$ , respectively); (C) construct integration: students enrolled in Biochemical Virology and Chemical Ecology significantly outperformed students in Plant Systematics ( $p = 0.0207$  and  $p = 0.0138$ , respectively) and in Environmental Restoration ( $p < 0.0001$  for both courses); (D) construct critical awareness: students in Chemical Ecology and Environmental Restoration scored significantly higher than students in Plant Systematics ( $p = 0.006$  and  $p = 0.016$ , respectively). The error bars represent the standard error of the mean.

participated in an interview (Table 3). Our first round of interview analyses were restricted to scoring the interviews binarily—students either articulated each construct as defined by the rubric (yes), or it was absent or scantily addressed (no).

*Evidence of Convergent Validity through Matched Data.* To better understand the data collected from the rubric, we compared same-student scores across essays and interviews ( $n = 25$ ;

Figure 4). Students' understanding of disciplinary grounding was relatively consistent across their essays and interviews, with 64% ( $n = 16$ ) of the population scoring high or low across both measurements. Overall, 11 students received high scores on their essays (apprentice to mastery) and exhibited this same level of understanding in their interviews; five students scored low (naïve to novice) on their essays, while also being unable to articulate disciplinary grounding in their interview (Figure 4A).



**FIGURE 4.** Numeric construct scores, (1) naïve, (2) novice, (3) apprentice, and (4) mastery, matched with same-student binary interview score (yes, no). (A) Disciplinary grounding, (B) integration, and (C) critical awareness. Bubble size corresponds to the number of students who obtained a given construct and interview score (i.e., larger bubbles indicate a greater number of students who received a particular matched score).

The construct integration had a much smaller proportion of matched understanding between essays and interviews ( $n = 11$ , 33%; Figure 4B). For critical awareness, the highest proportion of students ( $n = 17$ , 68%) had matched understanding between measurements (Figure 4C). However, of these 17 students, the majority ( $n = 13$ , 76%) scored low on critical awareness across both measurements, receiving between naïve (1) and novice (2) scores on their essays and binarily scored a “no” for their interviews. This accounts for more than half of the entire population (52%) who were not meeting the requirements for critical awareness set forth by the rubric across both measurements.

Many essay scores fell in between levels of understanding, (e.g., receiving a score of 3.5; see Table 4.A2). Reasons for this are twofold: 1) researchers were often unable to clearly identify if a student was exhibiting a mastery or apprentice, apprentice or novice, and/or novice or naïve level of understanding across the three constructs, and 2) averaging criteria scores often resulted in non-integers.

**Disciplinary Grounding.** Examples of students’ matched and mismatched scores for disciplinary grounding are provided in Table 4.A1. Chemical Ecology student Willow exhibits a mastery level of understanding in their essay and mirrors an adequate articulation of disciplinary grounding in their interview. Conversely, Birch from Chemical Ecology demonstrates a mismatched understanding of disciplinary grounding between their essay and interview. Birch’s essay exhibits high understanding of disciplinary knowledge (score of 3), but provides scant disciplinary methods (score of 1.5), thus receiving an average score of 2.25 for the construct. However, in Willow’s interview, they identify the exact methods to use when addressing a problem similar to the essay prompt.

**Integration.** The integration construct had the largest difference between essay and interview scores, with more than half of the population ( $n = 14$ , 56%) exhibiting integration knowledge in their interviews but unable to display this same understanding in their essays, receiving naïve or novice scores (Figure 4B).

Table 4.B1 demonstrates Plant Systematics student Cedar’s mastery level of understanding across both measurements in the construct integration, using two or more disciplines in an

integrated manner to advance the solution to the problem (meeting criteria requirements 3.1, 3.3, and 3.4). Cedar provides deep explanation of integrating biology (phylogenetic methods, using microsatellites, etc.) and chemistry (measuring alkaloids, glucosinolates) while connecting how each discipline is necessary. Similarly in Cedar’s interview, they describe how chemistry, biology, geology, and history are used in understanding each discipline, as well as how they build upon one another to yield a holistic understanding of the issue.

In Table 4.B2, Magnolia from Environmental Restoration displays a disconnected understanding of the construct integration. In this essay, no disciplines are included to support their approach to the essay prompt. Magnolia repeatedly poses questions throughout the essay but never attempts to provide an answer. But in the interview, they provide clear evidence of their understanding of integration through the identification of multiple disciplines and the connection between those fields (biogeochemistry, systems science, history) to yield a more well-rounded view of restoration, while relying on experts in other fields (hydrologist, geologist, biologist, chemist) to advance the solution toward a direction with the most successful outcome.

**Critical Awareness.** Student scores for the critical awareness construct conveyed a pervasive mismatched understanding across both measurements, with more than half of the population unable to display critical awareness as defined by the rubric in the essay or interview (Figure 4C).

In Table 4.C1, Maple from Plant Systematics received a mastery level score for critical awareness, including a description of benefits and limitations of integrating biology and chemistry methods and a metacognitive checkpoint for dealing with unexpected results. This level of awareness is paralleled in Maple’s interview as they discuss the all-angled thinking involved in biology research.

Hazel from Biochemical Virology, however, provides a grandiose outlook on the proposed solution to the essay prompt Table 4.C2. They provide no awareness of limitations and place extensive weight on benefits that are unfeasible given their approach in the essay. However, Hazel articulates an analytical critical awareness in their interview by explaining the limitations of research (experimental design, variables,

TABLE 4. Examples of matched and mismatched understanding of ID from same-student essay and interview responses

Construct	Essay responses	Interview quotes
Disciplinary grounding	<p style="text-align: center;"><b>A1. Willow, Chemical Ecology</b></p> <p>“The unknown plant bears fruits that appear healthy and edible, but without analysis of their nutritional content nothing can be said for certain. We intend on determining the mineral content of the fruit using near-infrared reflectance spectroscopy, as well as measuring secondary-metabolites to deter herbivory. Assessing floral morphology will provide insight into its pollination syndrome, and, consequently, its method of pollination.”</p> <p><b>Avg. construct score: mastery (4)</b></p>	<p>“I think about how plants use compounds, there’s all sorts of ecological relationships between plants, and different organisms, and pollinators, and the idea of plants producing nectar has a lot to do with chemistry. Then plants producing all sorts of volatile compounds that attract predatory organisms for defences.”</p> <p><b>Matched understanding: yes</b></p>
	<p style="text-align: center;"><b>A2. Birch, Chemical Ecology</b></p> <p>“The morphological character of the flower also does not indicate bee pollination. The inflorescence consists of a single yellow-orange tubular corolla with a deep nectar reserve, which suggests pollination by <i>Lepidoptera</i> or possibly hummingbirds. Further tests need to be conducted to figure out which one.”</p> <p><b>Avg. construct score: novice (2.25)</b></p>	<p>“We talked about compounds and secondary compounds of plants. There’s even, when you go down to systematics you’re talking about how things are related. To find out how things are related you look at the DNA of plants the molecular level through DNA sequencing and GenBank as well as they work morphologically.”</p> <p><b>Matched understanding: no</b></p>
Integration	<p style="text-align: center;"><b>B1. Cedar, Plant Systematics</b></p> <p>“We will perform a phylogenetic analysis using microsatellites to find out what species of fruit or vegetable this plant is most closely related to. We will use microsatellites since this new species must have recently diverged from an extant crop plant species. We can then contact chemists to analyze the chemical compounds present and correlate this with related species from the phylogenetic analysis.”</p> <p><b>Avg. construct score: mastery (4)</b></p>	<p>“It’s important to know how things are actually working, requiring the knowledge of chemistry and viewing biological systems in a chemistry sort of lens. Learning about geology and chemistry would really help in phylogenetic projects, just because understanding the history of the earth and the geography can help us interpret trends in the genotypes of organisms. The moulding of these knowledge sets ends in a greater understanding of plants holistically.”</p> <p><b>Matched understanding: yes</b></p>
	<p style="text-align: center;"><b>B2. Magnolia, Environmental Science</b></p> <p>“How the park will be restored mostly comes down to the project goals. This is a public park after all [...] not a far out wilderness ecosystem. So, what does the public want?”</p> <p><b>Avg. construct score: naïve (1)</b></p>	<p>“[Environmental restoration] means using systems science and science of cycles in biogeochemistry. It’s trying to bring back a previous state using history to look back at reference sites. Restoration requires collaborating between experts, having a more well-rounded view, because you’re bring[ing] in hydrologist to geologist, a biologist, a chemist. You’re thinking about all the different aspects of something instead of being one sided.”</p> <p><b>Matched understanding: no</b></p>
Critical awareness	<p style="text-align: center;"><b>C1. Maple, Plant Systematics</b></p> <p>“If the species is determined to be a self-pollinator and we determined the origin of its evolution through genetic sequencing there is a possibility that we could use cross pollination. However, as many self-pollinators use wind or rain as transportation modes for pollen, this could ultimately lead to an uncontrolled spread of the plants’ genes to other species, thus having a negative effect [on] the ecosystem. Alternatively, we could assess pollination through the measurement of volatile organic compounds. If all else fails, I would reassess my methodological approach.”</p> <p><b>Avg. construct score: mastery (4)</b></p>	<p>“I like the, ‘it may or may not happen this way’, in biology. I love going out into nature and [wondering], ‘Why is it that way?’ It is very important to set it up beforehand, like my bee pollination experimental design, and map it out and it may not go as planned. A big part of science is just recognizing why you failed or how you can do things better the next time around. Why didn’t they pollinate? Why did the plants not sprout? Why did we not get the results that we wanted? You need to go need back and check your experimental process!”</p> <p><b>Matched understanding: yes</b></p>
	<p style="text-align: center;"><b>C2. Hazel, Biochemical Virology</b></p> <p>“We can live in a better world, and this better world must inherently include all people on the planet earth. By providing a sustainable, high nutrient food source, we can [achieve] this dream thereby halting human starvation.”</p> <p><b>Avg. construct score: naïve (1)</b></p>	<p>“Learning about how to deal with experiments not turning out how you want them to turn out—what’s possibly good data when addressing the behemoth issue of food insecurity. Learning to take a step back—which variable or parameters are we going to change here to make this still useful, even though it didn’t turn out how we wanted it to turn out.”</p> <p><b>Matched understanding: no</b></p>

parameters) and the usefulness of alternative routes when an approach fails.

*The Emergent Theme of Collaboration.* Although there was no requirement to include a collaboration component in the essay submission or rubric, many students included language that indicated the necessity of collaboration. Of the larger population that completed the essay ( $n = 71$ ), 34% included the importance of collaborating with other scientists or community members in their essays; in the matched data set, 51% of students discussed elements of collaboration. Below is an example of collaboration language from a student essay:

“Regardless of the fire severity, including the public in the decision-making process should be a key component in the restoration program. The land is also built on indigenous grounds and it is critical to involve tribal members.”—Acacia, Environmental Restoration

### Research Question 3. How Do Undergraduate Students Perceive Interdisciplinary Science?

*Deductive Analysis.* Our first aim with the interview data was to code passages that aligned with constructs of the rubric, yet clearly students discussed ideas unrelated to the rubric. To capture these themes, we performed additional rounds of interview analysis using both inductive and deductive approaches. The codes that emerged from the initial inductive analysis mapped almost directly onto the subsequent, IDSF-derived, deductive codebook. Therefore, we chose to include the results and discussion from the deductive analysis only. This analysis allowed us to test the robustness of the IDSF criteria: disciplinary grounding, different research methods, integration, collaboration, and disciplinary humility. Examples of student interviews that reflected these criteria are outlined in this section.

*Perceptions of Disciplinary Grounding.* Disciplinary grounding is a shared construct between the rubric and IDSF. Of the 25 interview participants, 76% articulated disciplinary grounding:

“I did well [in this course] because of my larger knowledge in both chemistry and biology, like when we had to isolate cyanide. We had to specifically look at the plants and then we isolated the cyanide from various leaves. I guess our experimental process was a lot of biology. Then from there, we moved into the chemistry aspect. If I didn’t have both of those backgrounds, it would have made it hard for me to see the relevance or actually just get through the entire process.”—Elm, Chemical Ecology

“I think it would be a lot more helpful to learn disciplines by themselves in order to connect them. There’s always going to be some chemistry when you talk about biology and vice versa. It would be a lot more helpful to have a good background in those disciplines first.”—Hazel, Biochemical Virology

Interestingly, although many students understood the value of deep knowledge in one discipline, they often coupled this appreciation with a clause endorsing integration as the essential next steps:

“I think there are some benefits to learning a discipline by itself ... if you’re only a chemist, and you only focus on chemistry you can be a really good expert at that, but I think that it’s more important to also see how it connects to other fields. If someone is really into just researching DNA, and only doing that one thing, there is some benefit to that, like you’ll be the expert in that one specific thing. But if you want to have more relevance to the world it’s probably better to have some background of what else is going on.”—Hemlock, Environmental Restoration

“I think that there’s a reason that we make these arbitrary, or not so arbitrary distinctions between chemistry and biology, and physics. I think that they are so full of information, and concepts that it make sense to separate them, but it also makes equal sense to unify them, and to show that they’re not separate. That they are all the same system.”—Cedar, Plant Systematics

In expressing the value of disciplinary knowledge, many students exhibited the need for collaboration, and in doing so, were displaying disciplinary humility—an openness to and respect for other disciplines and value of collaboration in ID science:

“You can’t be an expert in everything. Depending on who you are, how you learn, what you’re passionate about, it may be better for some people to just focus on one discipline and they can become an expert in that. But they then should work with others with specialties in other areas to accomplish these heavier issues in society.”—Aspen, Plant Systematics

*Different Research Methods.* The inclusion of *different research methods* from multiple disciplines was overwhelmingly absent from student responses when describing ID science. Only 12% ( $n = 3$ ) of students included aspects of different disciplinary research methods in their conceptualization of ID science:

“Interdisciplinary science is combining chemistry, biology and ecology all together. I’m thinking, specifically, of tomato plants and the insects, the insects they interact with. They produce those alkaloids, which is a compound, a chemical compound. The caterpillars then, use that like a defense in an ecological system for predators.”—Willow, Chemical Ecology

*Advancement through Integration.* The idea of integration as integral to ID science was represented in all student interviews ( $n = 25$ , 100%) as they discussed the meaning of ID science:

“I think [interdisciplinary science is] kind of combining the different aspects of science, meaning that link between chemistry and ecology, biology. I think even physics can be thrown in there, and geology. Just kind of bringing it all together.”—Cherry, Plant Systematics

“I just think it’s important to know how things are actually working, which, a lot of the time, requires the knowledge of chemistry and viewing biological systems in a chemistry sort of lens. To get that full picture, you really need to look at the big thing.”—Sycamore, Biochemical Virology

Most students were not only juxtaposing multiple disciplines in their understanding (as displayed in the previous two quotes), but they often expressed integration through the *leveraging* of different knowledge and methods from separate disciplines to understand a phenomenon or advance knowledge (92%,  $n = 23$ )—a critical component that separates interdisciplinarity from cross- and multidisciplinary:

“Interdisciplinary science is important, like, learning geology would really help in phylogenetic projects, just because understanding the history of the earth and the geography can help us interpret trends in the genotypes of organisms.”—Spruce, Plant Systematics

“I think [interdisciplinary science] means the soft and hard sciences working together and building off each other’s knowledge. Understanding the human components. Bringing those together to understand how systems don’t work in a vacuum, and human components are kind of always at work in natural science systems.”—Magnolia, Environmental Restoration

“I really appreciate the interdisciplinary connections in science and I think that reflects a lot of true science. You might start out with one question but by the time you meet with other people who have knowledge in other regions, you may be able to ask more profound question and integrate your knowledge with their knowledge into the project. Integrating knowledge helps me learn how to deal with experiments not turning out how I want them to turn out, you know, how can we rethink this—what’s possibly good data. So learning to take a step back and lean on others [*sic*] methods and such.”—Larch, Chemical Ecology

**Integration through Collaboration.** Students spoke of collaborative efforts often in conjunction with integration language, as these two criteria are intricately entwined. When individuals collaborate, they bring their expertise to a team with hopes of successfully integrating pieces of knowledge with their collaborators to advance a field, create a discovery, or inform gaps in knowledge. This interconnection was expressed by 64% ( $n = 16$ ) of students:

“When you study something you know [it] really well, you do that one thing super well, but you may fail to take into account other factors that may be present or influencing it. You have to take it all into account and think about bigger pictures while at the same time looking at the small picture and the context. That’s hard for just one person to do when you’re thinking about a study which is why collaborating with a lot of different folks is important. I think that’s more important than focusing on details that aren’t seen in day-to-day life.”—Oak, Environmental Restoration

“It’s like in economic theory, this whole idea that if you have everybody doing everything, then you have a net loss. Meaning, if you have a person who’s a farmer and a doctor and trying to do everything at once, then you’re going to be much less productive in everything. Then if you have one person specializing as a farmer, and that’s all he does, and one person who’s a doctor and specialize[s] in that, then you can be more specialized in that field and you can share your information and everybody gains.”—Elm, Biochemical Virology

Many students touched on aspects of common ground (28%,  $n = 7$ )—a key contributor to successful collaborations as noted in the IDSF—and its importance in learning ID science:

“In this class, you never knew who you were going to end up talking to, where they were at as far as, like, conceptualizing what you are saying, or conceptualizing your project. So kind of having to adapt to that and make sure, you know you get so used to talking to people in your specific discipline, it was kind of nice to talk to other people and be like, oh that’s not even a thing in their world. Let me explain it. Or like same thing for me. I had to learn a lot about chemistry and different applications of that in biology.”—Pine, Chemical Ecology

“[Interdisciplinary science is] just trying to bridge the gap. Science is just trying to bring everything together, the whole because basically each discipline has similar things, but they’re from very different perspectives. So, if you can create a common language, it’d bring everybody together. You’d think it would be very beneficial, especially for medicine. If you’re trying to create drugs, you need to have crosstalk between different professors, etc., it’s how you bring that work together.”—Ash, Biochemical Virology

**Expression of Disciplinary Humility.** As we read and coded student interview transcripts, it was evident that students were expressing high levels of disciplinary humility (as defined in the IDSF) when verbalizing the meaning of ID science and its value in the larger context of society. Students communicated the idea of disciplinary humility similarly to how the IDSF defines this criterion—openness and respect for other disciplinary perspectives and expertise—throughout all of the aforementioned themes. We also identified interview responses that specifically capitalized on this humility (60%,  $n = 15$ ), and to a much lesser degree (12%,  $n = 3$ ), responses that connected the importance of leveraging STEM and non-STEM disciplines:

“I don’t think that you can have one [discipline] without the other when you’re talking about any type of science and that includes social sciences. Unless you want to just stick yourself in the lab all day and never talk to anyone else, which is totally fine, but you’re going to have to know what your research is doing and how it connects with others’ research in order to, kind of, elevate the importance.”—Olive, Environmental Restoration

“I think we lose a lot of knowledge when we ignore that someone else might have a different way to interpret things especially given their background. For instance, I think it is important to get different data interpretations. Everybody has a different way to interpret things.”—Juniper, Chemical Ecology

I think [interdisciplinary science] can help further research and improve it, and also help solve real world science problems. I think with restoration ecology in particular, you need combination of different scientists, including those from the soft sciences, so if they already have that knowledge of other fields, it will improve their problem solving abilities.”—Sassafras, Environmental Restoration

## DISCUSSION

The necessity of ID science as a critical factor in solving real-world problems is undeniable. Yet, little has been done to assess

whether future scientists are equipped with the resources to address this competency at the undergraduate level. Here, we identified that instructors typically assess ID science understanding through writing assignments, and therefore developed an essay assignment as a platform for students to exhibit their ID science knowledge. We then tested for evidence of convergent validity of data collected from a preexisting ID rubric to evaluate undergraduate students' understanding of this competency. Finally, we used our results to inform the robustness of the IDSF through a similar validity analysis.

### Instructors Assess ID Science Understanding through Writing

The faculty we surveyed predominantly identified writing activities as the main way they assess ID understanding in a classroom setting (69%; Table 2). Within this category, more than half of our participants identified writing assignments through essays and papers as the preferred method of assessment (51%). This finding is consistent with a wealth of literature that suggests students must be given the opportunity to problem solve and think critically through writing when addressing the complexities involved in ID science (Boix-Mansilla *et al.*, 2009; Chan *et al.* 2010; Balgopal *et al.*, 2012, 2017; Gouvea *et al.*, 2013; Cooper and Stowe, 2018; Tripp and Shortlidge, 2019). Thus, we support that instructors use writing-intensive activities as a mechanism to measure ID science understanding in an undergraduate setting.

Traditional assessments (34%) through exams and quizzes were chosen as the second most frequently used evaluation, followed by group work (34%). Although group work is being championed through active-learning models (Johnson and Johnson, 2009; Haak *et al.*, 2011; Lamm *et al.*, 2012; Wilson *et al.*, 2018), and, importantly, is how ID work is actually accomplished, this was not well reflected by science faculty. Such results suggest that experts rely on students being able to demonstrate that they can think interdisciplinarily before actually participating in a collaborative ID project (Tripp and Shortlidge, 2019). We recommend that, as we consciously move students through a progression from thinking to acting interdisciplinarily, we should thoughtfully consider the appropriateness of the assessment method.

### The Rubric Can Differentiate Performance Based on Constructs among Students and Courses

The rubric was able to detect differences in student construct performance based on the rubric scores from the essay (Figures 1–3) as defined by the designers (Boix Mansilla *et al.*, 2009). Overall, students struggled to meet the requirements for integration and critical awareness understanding, as evidenced by the significantly lower scores in these constructs compared with purposefulness and disciplinary grounding across our entire population (Figure 1). This is unsurprising, as integration and critical awareness are much more nebulous and are not well defined for the natural and physical sciences (Borrego *et al.*, 2009). These statistical analyses support findings from the matched data of essay scores and interviews—students are conceptualizing disciplinary grounding similarly to the rubric and thus have higher scores in this construct, reflecting this understanding. The significantly lower student scores for integration and critical awareness are also reflected in the matched data—

students are operationalizing these constructs differently than the rubric and perhaps more similarly to the IDSF. Although there were no matched data for purposefulness, students scored well on this construct, providing evidence that this may be an important piece in helping student to frame their writing in an appropriate context.

When comparing differences in overall student performance in essays by course and rubric constructs, the students enrolled in Chemical Ecology, which was a course-based undergraduate research experience, had higher scores than students in Biochemical Virology, Environmental Restoration, and Plant Systematics across all rubric constructs (Figures 2 and 3). This high performance may be a result of students exposure to ID science as they worked on a chemical ecology research question with a biologist and a chemist. Students taking a research-based course have the chance to “do” science as opposed to learn about science, which in practice is often an inherently ID endeavor. Additionally, students enrolled in courses taught by two instructors from different disciplines (Chemical Ecology and Biochemical Virology; see Table 3 for information on instructors) tended to score higher on total average essay scores than students taught by one disciplinary instructor (unpublished data). This may reinforce that students need exposure to deep disciplinary knowledge from separate disciplines to effectively integrate knowledge to address real-world issues.

### The Rubric Did Not Fully Measure Up to Validity Tests across Data Measurements

Our matched data between essays and interviews reveal several inconsistencies between students' written and verbal communication of the integration and critical awareness constructs, while fewer inconsistencies arise in disciplinary grounding. There was variance across students' matched scores within the disciplinary grounding construct (Figure 4A), however; the construct appears to be operating well. Individual students received similar scores in both their interview and essay measurements (e.g., same-student high scores on interviews and essays or same-student low scores across both). In interviews, students also expressed an appreciation for disciplinary knowledge, which supports the IDSF's disciplinary grounding category and provides further evidence that this construct is likely a fundamental component to ID science understanding.

Integration (Figure 4B) had the largest amount of mismatched understanding between same-student measurements, with many students meeting this benchmark in their interviews but completely missing this understanding in their essays ( $n = 12$ , measured as scores below 2.5 in the rubric). This could be attributed to certain criteria within this construct being unsuitable for our population. If students are able to articulate integration in interviews but miss this mark entirely in their written responses, this could suggest that natural and physical science students may operationalize integration differently from the social sciences. This point is further substantiated in our deductive analysis of interviews using the IDSF—students are indeed understanding integration through the leveraging of disciplines to advance solutions to problems. Thus, it is likely that elements of the rubric are either stifling student incorporation of integration or leading them toward more simplified connections between disciplines. We also fully recognize that students'

abilities vary between written and oral assessments, perhaps partially contributing to this discrepancy. However, it is unlikely that this is the case for more than half of our population. Redefining and setting clear expectations for accomplishing integration as defined in the IDSF could potentially provide an outlet for students to apply their integrative knowledge to real-world issues.

Critical awareness (Figure 4C) scores were perhaps the most perplexing of the constructs, with more than half of our population completely missing or poorly meeting this construct across both essays and interviews ( $n = 13$ , measured as scores below 2.5 in the rubric). A potential reason for this may be similar to issues surrounding the integration construct—the operationalization of critical awareness may differ between the natural and social sciences (Borrego *et al.*, 2009). Another possibility is that critical awareness may be beyond a student's capability at the undergraduate level, or perhaps this construct is not suitable for the natural and physical sciences altogether. We posit that students should indeed be critically aware of the strengths and limitations of a study or approach, but perhaps should not be expected to be proficient in recognizing all of the caveats of each discipline involved, as the rubric would suggest. Instead, it may be highly beneficial to restructure critical awareness to more closely align with the IDSF's criteria for disciplinary humility. Students overwhelmingly expressed an awareness of, and respect for, expertise and perspectives from other disciplines during their interviews, frequently expressing their lack of knowledge in other disciplinary domains. This humility and consciousness of their own limitations is critical in ID science work and are characteristics of conscientious scientists. To nurture this humble mindset on a larger scale, students may need to be led through a more explicit pathway, such as the IDSF, to develop a greater ability to demonstrate open-mindedness and inclusivity in ID collaborations (particularly with non-STEM disciplines, as discussed in the next section).

Furthermore, we propose that a lack of understanding of both integration and critical awareness on the part of students may not be their fault, but reflects a need for instructors to be more intentional in helping students integrate knowledge, concepts, and methods/methodologies across disciplines and provide opportunities for students to enhance their critical awareness by thinking “outside the box.”

### Exclusion of Non-STEM Disciplines

A common trend in the matched interview and writing assignment data sets was the absence of non-STEM disciplines in students' essay responses. According to the IDSF, disciplinary humility is the thread that connects all other aspects of ID science understanding with a component to consider non-STEM contributions to real-world problems (Tripp and Shortlidge, 2019). We developed essay prompts that we believed would elicit students integrating the humanities and social sciences (i.e., topics included declining honeybee populations, environmental restoration, and infectious viral outbreaks) for a complete, thoughtful response. However, only 25% of students mentioned disciplines outside STEM fields in their essays. The rubric may not have the necessary elements or may be too specific for students to consider including fields outside STEM. Further, specific courses had a stricter focus on

STEM, and thus it is possible that other fields were overlooked or students perceived that non-STEM disciplines would not be appropriate for these writing activities. This pervasive lack of non-STEM inclusion was also reflected in student interviews, with only 12% of students speaking to the importance of non-STEM disciplines. This is alarming, as mitigating real-world issues such as food insecurity will undoubtedly involve non-STEM fields. In addition, research suggests that undergraduate STEM students have historically been less mindful of societal issues and how science can impact equity and the human good (Garibay, 2015). To train STEM students to be more civically responsible and socially aware of the impact of science, instructors should make intentional efforts to incorporate connections between science and society into curricula and assessments (NRC, 2009; AAAS, 2011; Garibay, 2015).

### The Importance of Collaboration

Finally, the theme of collaboration emerged in both essays and interviews, with percentages appreciably high considering there was no specific collaboration requirement in the rubric (34% in the entire essay data set, 64% in interviews, 51% across the matched data set). These findings corroborate Borrego and Newswander's work (2010), stating that the natural and physical sciences rely on collaboration at higher rates than other disciplines. The inclusion of this theme is also strongly supported by the IDSF (Tripp and Shortlidge, 2019) highlighting the importance of interacting across disciplines and suggesting this as a fundamental cornerstone of ID science understanding.

### Students Conceptualize Interdisciplinary Science in Ways That Align with the IDSF

The majority of students discussed ideas that closely align with our previous work on science faculty's perceptions of ID science and the IDSF (Tripp and Shortlidge, 2019). Students' perceptions of ID science reflect almost all of the criteria in the IDSF as evidenced through interviews in this study. Many students exhibited the essence of disciplinary humility, acknowledging and respecting the importance of other disciplines within STEM during interviews (60%). Moreover, students often described attributes of ID science as the application of these different disciplines to solve larger societal problems. This result closely ties to elements of faculty's description of ID science and supports the notion that real-world problems inherently require the application of multiple disciplines in order to deal with the problem at hand (Tripp and Shortlidge, 2019). Students spoke to the importance of being grounded in disciplinary knowledge before integrating different disciplines (74%), often describing integration as the leveraging of disciplinary knowledge, methods, or ideas into a cohesive whole (92%). The majority of students included collaboration as a hallmark of ID science (64%) and, to a lesser extent, the necessity of common ground and/or language among ID collaborators (28%). This idea of common ground in ID work is emphasized by Tripp and Shortlidge (2019), as well as many ID experts, as a necessary component for effective ID collaboration (Boix Mansilla and Gardner, 2003; Thompson Klein, 2005; Öberg, 2009).

One pillar of understanding ID science according to the IDSF that was severely lacking recognition in student

interviews is the inclusion of “different research methods” from other disciplines. However, students effectively met a similar criterion (2.2: Does the paper use disciplinary methods accurately and effectively (e.g., experimental design?) in their essays, as evidenced by their high scores in disciplinary grounding. Undergraduate students’ inexperience with research possibly explains this exclusion in their interviews. We contend that, as students start to engage in more ID science research and collaborations, this awareness of different research methods will likely increase.

We believe that the alignment of student perceptions and conceptualization of ID science with the IDSF increase the validity of the IDSF as an accurate model to design curricula that capture students’ understanding of ID science.

### Limitations

We acknowledge that we used only one non-ID course, resulting in the absence of statistical analyses of student scores based on ID course format (ID vs. non-ID), difference in demographics, and prior ID science exposure. Trends in our data indicated that, rather than focus on differences between ID and non-ID course format, individual student scores based on demographics, or prior ID experience, efforts should first be centralized around developing a functional tool that effectively captures student understanding of ID science across disciplines on a larger scale, regardless of student background or classroom environment. Although our effect sizes were large, the significant differences that were observed in scores between the research-based course and the other courses represent a small sample size, and more essays should be analyzed from multiple research- and non-research-based courses to provide further evidence on the instrument’s functionality. We also recognize that these data were collected from students from only one institution across four upper-division courses and thus may not be representative of other student populations.

Additionally, we did not prompt students in interviews to specifically verbalize ideas related to each rubric construct, as we wanted another means of gathering students’ understanding and perceptions of ID science outside the rubric—this allowed us to gain evidence of convergent validity. However, because of this, our interview results may not be fully inclusive of student perceptions and knowledge specifically related to constructs defined in the rubric. Finally, the rubric constructs were not necessarily linked to learning outcomes and/or focal points of course material, as we did not consult with instructors about embedding the constructs or related ideas into the course before the assignment.

### CONCLUSION AND NEXT STEPS

Conceptualizing how ID science understanding is currently met and finding ways to assess undergraduate students’ ID understanding is imperative if we are to meet the ID expectations set forth by initiatives such as *Vision and Change*. The development of writing activities is one potential platform for students to express their understanding of ID science in a creative yet constructive way. Providing evidence of validity of data collected from preexisting instruments and frameworks can inform the selection and/or development of instruments to assess ID science understanding through these activities.

The results of this study do not provide sufficient evidence that valid data were collected using the preexisting rubric, yet they largely support the criteria outlined in the IDSF (disciplinary grounding, different research methods, integration, collaboration, and disciplinary humility). We aim to develop and test an instrument based on factors that functioned well with the rubric and constructs that align with the IDSF. Future efforts will focus on gathering a larger sample of essay responses across student populations and conducting student/faculty interviews to further develop an instrument that provides valid data on students’ understanding of ID science. This research is a step toward being able to use best practices in measuring undergraduate science students’ ability to “tap into the interdisciplinary nature of science” as described by *Vision and Change* (AAAS, 2011).

### ACKNOWLEDGMENTS

We thank the instructors who worked with us to embed the essay in their classes and the student participants for their generous contributions and participation in this study. We thank undergraduate researchers Analee Pham and Megan Thran for their help with data analysis. We also thank the Biology and Chemistry Education Research Group at Portland State University for providing thoughtful feedback on interview questions.

### REFERENCES

- American Association for the Advancement of Science. (2011). *Vision and change: A call to action. Final report*. Washington, DC. Retrieved July 22, 2019, from <http://visionandchange.org/finalreport>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2013). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Balgopal, M. M., & Wallace, A. M. (2009). Decisions and dilemmas: Using writing to learn activities to increase ecological literacy. *Journal of Environmental Education*, 40(3), 13–26.
- Balgopal, M. M., Wallace, A. M., & Dahlberg, S. (2012). Writing to learn ecology: A study of three populations of college students. *Environmental Education Research*, 18(1), 67–90.
- Balgopal, M. M., Wallace, A. M., & Dahlberg, S. (2017). Writing from different cultural contexts: How college students frame an environmental SSL through written arguments. *Journal of Research in Science Teaching*, 54(2), 195–218.
- Barbera, J., & VandenPlas, J. R. (2011). All assessment materials are not created equal: The myths about instrument development, validity, and reliability. In Bunce, D.M. (Ed.), *Investigating classroom myths through research on teaching and learning* (Vol. 1074, pp. 177–193). Washington, DC: American Chemical Society. <https://doi.org/doi:10.1021/bk-2011-1074.ch011>
- Besterfield-Sacre, M., Gerchak, J., Lyons, M. R., Shuman, L. J., & Wolfe, H. (2004). Scoring concept maps: An integrated rubric for assessing engineering education. *Journal of Engineering Education*, 93(2), 105–115. <https://doi.org/10.1002/j.2168-9830.2004.tb00795.x>
- Boix Mansilla, V., & Duraisingh, E. D. (2007). Targeted assessment of students’ interdisciplinary work: An empirically grounded framework proposed. *Journal of Higher Education*, 78(2), 215–237.
- Boix Mansilla, V., Duraisingh, E. D., Wolfe, C. R., & Haynes, C. (2009). Targeted assessment rubric: An empirically grounded rubric for interdisciplinary writing. *Journal of Higher Education*, 80(3), 334–353. doi: 10.1353/jhe.0.0044
- Boix Mansilla, V., & Gardner, H. (2003). Assessing interdisciplinary work at the frontier. An empirical exploration of “symptoms of quality.” *Interdisciplinary Studies Project, Project Zero*. Harvard Graduate School of Education Publications.

- Borrego, M., & Newswander, L. K. (2010). Definitions of interdisciplinary research: Toward graduate-level interdisciplinary learning outcomes. *Review of Higher Education, 34*(1), 61–84.
- Borrego, M., Newswander, C. B., McNair, L. D., McGinnis, S., & Paretto, M. C. (2009). Using concept maps to assess interdisciplinary integration of green engineering knowledge. *Advances in Engineering Education, 1*(3), n3.
- Carlson, C. A. (2007). A simple approach to improving student writing. *Journal of College Science Teaching, 36*(6), 48.
- Chan, Y.-Y., Yu, A. C. H., & Chan, C. K. K. (2010). Assessing students' integrative learning in biomedical engineering from the perspectives of structure, behavior, and function. In *2010 IEEE Frontiers in Education Conference (FIE)* (pp. S1G-1–S1G-6). Washington, DC: IEEE. <https://doi.org/10.1109/FIE.2010.5673660>
- Connolly, P., & Vilardi, T. (1989). *Writing to learn mathematics and science*. New York: Teachers College Press.
- Cooper, M. M., & Stowe, R. L. (2018). Chemistry education research—From personal empiricism to evidence, theory, and informed practice. *Chemical Reviews, 118*(12), 6053–6087. <https://doi.org/10.1021/acs.chemrev.8b00020>
- Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report, 20*(9), 1408.
- Garibay, J. C. (2015). STEM students' social agency and views on working for social change: Are STEM disciplines developing socially and civically responsible students? *Journal of Research in Science Teaching, 52*(5), 610–632. <https://doi.org/10.1002/tea.21203>
- Gouvea, J. S., Sawtelle, V., Geller, B. D., & Turpen, C. (2013). A framework for analyzing interdisciplinary tasks: Implications for student learning and curricular design. *CBE—Life Sciences Education, 12*(2), 187–205.
- Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science, 332*(6034), 1213–1216.
- Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*, 365–379.
- Keys, C. W. (1999). Revitalizing instruction in scientific genres: Connecting knowledge production with writing to learn in science. *Science Education, 83*(2), 115–130.
- Lamm, A. J., Shoulders, C., Roberts, T. G., Irani, T. A., Snyder, J. L., & Brendemuhl, B. J. (2012). The influence of cognitive diversity on group problem solving strategy. *Journal of Agricultural Education, 53*, 18–30.
- Maher, J. M., Markey, J. C., & Ebert-May, D. (2013). The other half of the story: Effect size analysis in quantitative research. *CBE—Life Sciences Education, 12*(3), 345–351. <https://doi.org/10.1187/cbe.13-04-0082>
- National Research Council (NRC). (2003). *BIO2010: Transforming undergraduate education for future research biologists*. Washington, DC: National Academies Press.
- NRC. (2009). *A new biology for the 21st century*. Washington, DC: National Academies Press.
- Öberg, G. (2009). Facilitating interdisciplinary work: Using quality assessment to create common ground. *Higher Education, 57*(4), 405–415.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, CA: Sage.
- President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Washington, DC: U.S. Government Office of Science and Technology.
- Rivard, L. O. P. (1994). A review of writing to learn in science: Implications for practice and research. *Journal of Research in Science Teaching, 31*(9), 969–983.
- R Studio Team. (2019). *RStudio: Integrated development for R*. Boston, MA: RStudio. Retrieved August 14, 2019, from [www.rstudio.com](http://www.rstudio.com)
- Saldaña, J. (2015). *The coding manual for qualitative researchers*. Newbury Park, CA: Sage.
- Stangor, C. (2014). *Research methods for the behavioral sciences* (5th ed.). New York: Houghton Mifflin.
- Thompson Klein, J. (2005). Interdisciplinary team work. The dynamics of collaboration and integration. In Derry, S., Schunn, J. C., & Gernsbacher, D. M. A. (Eds.), *Interdisciplinary collaboration. An emerging cognitive science* (pp. 23–50). Mahwah, NJ: Erlbaum.
- Tripp, B., & Shortlidge, E. E. (2019). A framework to guide undergraduate education in interdisciplinary science. *CBE—Life Sciences Education, 18*(2), es3. <https://doi.org/10.1187/cbe.18-11-0226>
- Wilson, K. J., Brickman, P., & Brame, C. J. (2018). Group work. *CBE—Life Sciences Education, 17*(1), fe1.